Making Reliability Engineering Smart: When Principles of Failure Meet with Industrial Big Data

Zhiguo Zeng

université PARIS-SACLAY

CentraleSupélec

NNT : 20XXSACLXXXX

Thèse de HDR

Making Reliability Engineering Smart: When Principles of Failure Meet with Industrial Big Data

Thèse de l'habilitation à diriger des recherches (HDR) de l'Université Paris-Saclay
préparée à Centralesupélec

École doctorale n°ED 573 Interface: approches interdisciplinaires, fondements, applications et innovation (INTERFACES)
Spécialité de HDR: Ingénierie des Systèmes Complexes

Thèse présentée et soutenue à Gif-sur-yvette, le 13/07/2022, par

ZHIGUO ZENG

Composition du Jury :

JOUINI, Oualid Professeur (HDR), Centralesupélec, Université Paris-Saclay, France	Présidente, Examinateur
BASTIDAS-ARTEAGA Emilio Professeur (HDR), La Rochelle University, France	Rapporteur
MEDJAHER, Kamal Professeur (HDR), INPT-ENIT, University of Toulouse, France	Rapporteur
SCHOEFS, Franck Professeur (HDR), Université de Nantes, France	Rapporteur
COIT, David Professeur, Rugters University, USA	Examinateur

This thesis is dedicated to my parents,

Mr. Qifo Zeng and Honglian Wu,

for raising me up and helping me become what I am today;

to my wife, Dr. Shijia Du,

for tolerating and supporting me all the way to here;

and to my son, Zhuochen (Duke) Zeng,

for stopping crying occasionally and allowing me finish this work.

谨以这本论文，

献给我的父亲曾齐佛先生和母亲吴红莲女士，

感谢他们的养育之恩；

献给我的亲爱的妻子杜时佳博士，

感谢她一路走来对我的理解与支持；

也献给我们的宝宝曾卓宸，

感谢他偶尔停止哭闹，让我能挤出时间完成这篇论文。

Contents

List of Figures

List of Tables

Chapter 1

INTRODUCTION

I still remembered that about ten years ago, when I just started my PhD at Beihang University, I attended a seminar of a world-renowned expert in reliability, Prof. Way Kuo from City University of Hongkong, in which Prof. Kuo discussed main challenges in reliability research and gave his perspectives on the future of reliability engineering. He ended his seminar with a discussion on different reliability approaches V.S. their required data size. The very last comment he made was, "what if we have only one data point? (Making good reliability assessment under this circumstance) It is a vision." It was an excellent play on words, as the seminar was held in Vision Hotel: everyone left with a knowing smile on the face. However, it also proposes an important challenge in modern reliability engineering: if we do not have enough historical failure data, how can we still evaluate the reliability with sufficient degree of confidence?

The vision of Prof. Kuo has become one of the central issue of my research till today. In this thesis, my major results related to this topic are presented. The ultimate goal of these research activities is to improve the performance of reliability assessment and decision-making under practical constraints of lacking enough historical failure data. As the title of this thesis reveals, my research activities attempt to tackle this challenge from two angles. The first branch of research aims at understanding and modeling the failure behaviors from a physics-based perspective, and estimating the reliability based on the physics-based models. By accurately modeling the failure behaviors physically, the reliability can be estimated with good confidence, even though few historical data are available. Dependencies among the failure mechanisms and treatment of uncertainty are two issues that need special attention when developing the failure behavior models. Another branches of my research activities focus on using industrial big data, especially the online collected data during operation, to make up-to-date reliability assessment and remaining useful life prediction. Different types of data can be used, *e.g.,* condition-monitoring data, inspection data, expert judgment. The challenge here is that different data sources are often heterogeneous in nature. How to integrate the heterogeneous data sources needs investigation.

More specifically, the different research activities discussed in this thesis fall into five research axes, as shown in Figure 1.1. The purpose of the first axis is to develop a conceptual framework for understanding different contribut-

ing factors to failures. A detailed presentation can be found in Chap. 6. The developed conceptual framework will serve as theoretical foundations for the modeling and analyzing of failure behaviors. Degradation is an important contributing factor to failures in the developed conceptual framework. The second research axis, then, focuses on developing models and efficient assessment methods for dependent degradation processes. In particular, we focus on the degradation process that involve both continuous degradation and discrete state transitions (Chap. 7). The third research axis focuses on another important contributing factor to failures in the conceptual framework, *i.e.,* epistemic uncertainty. New methods are proposed for the practical evaluation of epistemic uncertainty and integration of epistemic uncertainty with the result of risk/reliability assessment (see Chap. 8 for a detailed presentation). The first three research axes focus on system failure behavior, without considering the potential performance recovery process. In the fourth research axis, we discuss how to model the behavior of a multi-state system whose performance can be recovered after initial failure or performance disruptions through a new modeling and analysis framework based on Markov/semi-Markov reward process is developed (see Chap. 9). Finally, in the last research axis (see a detailed discussion in Chap. 10), we discuss how to fuse different available data sources for online reliability assessment and remaining useful life prediction. In this axis, we also intend to investigate how to merge knowledge on physics of failure and the online collected data for better reliability assessment, as suggested in the title of this thesis.

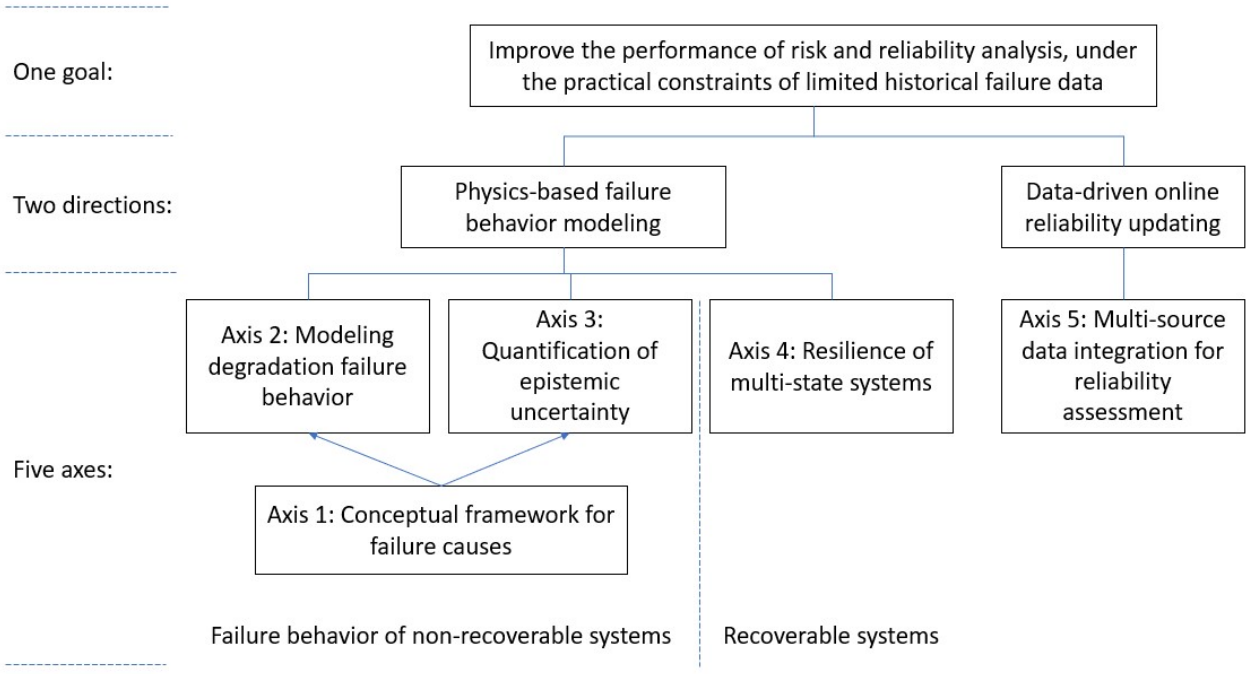


Figure 1.1: Structure of the research activities.

The research activities included are those conducted after my PhD defense (*i.e.,* from Jan. 2016). It serves as the main document supporting my application for the Habilitation à Diriger des Recherches (HDR) at Unveristé Paris Saclay. The thesis is organized in two parts. Part I (Chaps. 2 - 5) is mainly for administrative purposes: it presents

a synthetic summary of my research, teaching and student supervision activities that supports my application of HDR. In Chap. 2, my curriculum vitae is presented. Then, in Chap. 3, my activities related to student supervision and teaching are summarized. Chapter 4 is dedicated to synthetically present the research activities, including my awarded projects/grants, and community recognition of my research. Finally, in Chapter 5, a complete list of my publications is given. Part II (Chaps. 6 - 11) summarizes some of my major research results in the past and also briefly introduce my scientific project for the future. The presentation of the research results is organized in the five axes discussed before, in Chapters 6 to 10, respectively. The writing of these chapters are based on previous papers published by me and the PhD students I co-supervised. Chapter 11 presents the scientific project for my future research. Finally, the thesis is concluded in Chapter 12.

Part I

SYNTHETIC SUMMARY OF RESEARCH AND SUPERVISION ACTIVITIES

Chapter 2

CURRICULUM VITAE

PERSONAL INFORMATION

- Zhiguo ZENG (曾志国), born in 03/05/1989, in Fuzhou, Fujian, China.
- Current position: Assistant Professor, Chaire on Risk and Resilience of Complex Systems, Laboratoire Genie Industriel, Centralesupélec, Université Paris-Saclay, France.
- Nationality: Chinese.
- Languages spoken: Chinese (Native), English (Fluent (TOFEL iBT: 108/120)), French (Basic).

EDUCATION BACKGROUND

Ph.D. in Reliability and Systems Engineering Sep. 2011 - Dec. 2015

Beihang University Beijing, China Adviser: Prof. Rui Kang, Prof. Yunxia Chen

Dissertation Title: Belief Reliability Theory and Application: Measuring Reliability Under Influence of Epistemic Uncertainty.

Jury Members: Prof. Daqing Li, Prof. Xiang, Li, Prof. Shaoping Wang, Prof. Daoping Wang, Prof. Yongli Yu (chair).

Defense date: 18/12/2015 Location: Beijing, China.

Visiting Ph.D. Student Sep. 2015 - Dec. 2015

Politecnico di Milano Milano, Italy Adviser: Prof. Enrico Zio

B.Eng. in Quality and Reliability Engineering (GPA top 5%) Sep. 2007 - July, 2011

Beihang University Beijing, China

WORKING EXPERIENCES

Assistant Professor Dec. 2017 - Present

Centralesupélec, Université Paris-Saclay Paris, France

Research focuses: Modeling, Simulation and Optimization of Complex System and Critical Infrastructures

Postdoc Researcher April 2016 - Nov. 2017

Centralesupélec, Université Paris-Saclay Paris, France Adviser: Prof. Enrico Zio

Research focuses: Modeling, Simulation and Optimization of Complex System and Critical Infrastructures

Research Associate Jan. 2016 - March 2016

Beihang University Beijing, China Adviser: Prof. Rui Kang

Research focuses: Belief reliability modeling and analysis for complex engineering systems

RESEARCH INTERESTS

- Reliability Modeling, Simulation and Optimization for Complex System of industry 4.0 (New direction after my PhD),
- Modeling of Degradation Processes and Dependent Failure Behaviors (New direction after my PhD),
- Uncertainty Analysis (Particular: Epistemic Uncertainty. This is a continuation of my PhD thesis).

Chapter 3

SYNTHETIC PRESENTATION OF STUDENT SUPERVISION AND TEACHING

This chapter summarizes my student supervision and teaching activities from 2016 to present. Following the guidelines from Université Paris-Saclay for applying Habilitation à Diriger des Recherches (HDR) [1], this chapter is organized into five sections. Section 3.1 presents an executive summary of my student supervision activities, including both PhD and master students. Section 3.2 details my supervision activity for each student, together with a list of co-authored papers with the students. Section 3.3 presents my other activities related to PhD training. Section 3.4 presents a summary of my teaching activities. Section 3.5 presents detailed information of the courses I have taught. These sections are prepared to demonstrate the my compliance to the criteria for obtaining HDR at Université Paris-Saclay, defined in Sect. 1.3 of [4].

3.1 STUDENT SUPERVISION: SUMMARY

- PhD students: 4 completed and 6 on-going.
- Master students: 3 completed and 3 on-going.
- Awards received by supervised/co-supervised students:
  - Miss Mengfei Fan, PhD student, **"Outstanding PhD thesis award"**, Beihang university, 2018.
  - Miss Taneem Bani-Mustafa, PhD student, **"Best presentation award"**, ICSRS 2018.
  - Mr. Qingyuan Zhang, PhD student, **"Excellence Research Grant for Outstanding PhD Students"** from Beihang university.

– Mr. Raghed Saab, Master student, **"International Internship Travel grant"**, French embassy in Libanon, 2020.

3.2 COMPLETE LIST OF SUPERVISED STUDENTS

3.2.1 PhD students: 4 completed and 6 on-going

- Miss Jinduo Xing (Centralesupélec, France, Sep. 2015 - Dec. 2019):
  - Thesis title: Business continuity of energy systems: A quantitative framework for dynamic assessment and optimization.
  - Co-supervised (50%) with Prof. Enrico Zio (50%).
  - Graduated with 3 publications in international journals, 2 in international conferences.
  - Current position: Assistant professor, Beijing University of Architecture and Construction.
  - Selected publications with me:
    1. Xing J., **Zeng Z.***, Zio E., Joint optimization of safety barriers for enhancing business continuity of nuclear power plants against steam generator tube ruptures accidents. Reliability Engineering and Systems Safety. 2020; 202, 107067. (JCR Q1).
    2. Xing J, **Zeng Z***, Zio E. Dynamic business continuity assessment using condition monitoring data. International Journal of Disaster Risk Reduction 2019, 41, 101334. (JCR Q2).
    3. Xing J, **Zeng Z***, Zio E. A framework for dynamic risk assessment with condition monitoring data and inspection data. Reliability Engineering and Systems Safety 2019, 191, 106552. (JCR Q1).
    4. Xing J, **Zeng Z** and Zio E. An integrated framework for condition-informed probabilistic risk assessment. Proceedings of Annual European Safety and Reliability Conference (ESREL2017), Portoroz, Slovenia, 2017.
- Miss Tasneem Bani-Mustafa (Centralesupélec, France, Sep. 2015 - Dec. 2019):
  - Thesis title: Multi-Hazards Risk Aggregation Considering the Trustworthiness of the Risk Assessment.
  - Co-supervised (50%) with Prof. Enrico Zio (50%).
  - Graduated with 4 publications in international journals, 3 in international conferences.
  - Awarded "**best presentation award"** in ICSRS 2018.
  - Current position: Chef du projet, Nuclear consulting services.
  - Selected publications with me:

1. **Zeng Z**, Bani-Mustafa T, Flage R, Zio E. An integrated risk index accounting for epistemic uncertainty in Probability Risk Assessment (PRA). Journal of Risk and Reliability 2020. (JCR Q3).
2. Bani-Mustafa T, Flage R, **Zeng Z**, Zio E. An extended method for evaluating assumptions deviations in quantitative risk assessment and application to external flooding risk assessment of a nuclear power plant. Reliability Engineering and Systems Safety 2020; 200, 106947. (JCR Q1).
3. Bani-Mustafa T, **Zeng Z**, Zio E, Vasseur D, A practical approach for the evaluation of the strength of knowledge supporting risk assessment models. Safety Science 2020; 124, 104596. (JCR Q1).
4. Bani-Mustafa T, **Zeng Z**, Zio E, Vasseur D, A new framework for multi-hazards risk aggregation. Safety Science 2020; 121, 283-302. (JCR Q1).
5. Bani-Mustafa T, Zeng Z, Zio E and Vasseur D, A Framework for Multi-Hazards Risk Aggregation Considering Risk Model Maturity Levels, ICSRS2017, Milano, 2017.
6. Bani-Mustafa T, **Zeng Z**, Zio E, Vasseur D, Strength of Knowledge Assessment for Risk Informed Decision Making. Proceedings of Annual European Safety and Reliability Conference (ESREL2018), Trondheim, Norway, 2018.

- Miss Mengfei Fan (Beihang University, China, Sep. 2015 - Dec. 2018):
  - Thesis title: Stochastic hybrid system-based modelling of dependent failure processes.
  - Co-supervised (50%) with Prof. Rui Kang (50%).
  - Graduated with 5 publications in international journals, 4 in international conferences.
  - Recipient of **"Outstanding PhD thesis award"** from Beihang university.
  - Current position: Senior engineer, The Second Institute of China Aerospace Science and Technology Corporation.
  - Selected publications with me:
    1. Zio E, Fan M, **Zeng Z***, Kang R, Application of reliability technologies in civil aviation: lessons learnt and perspectives. Chinese Journal of Aeronautics. 2019 (32) 1: 143-158. (JCR Q1).
    2. Fan M, **Zeng Z***, Kang R, Zio E and Chen Y. A Sequential Bayesian Approach for Remaining Useful Life Prediction of Dependent Competing Failure Processes. IEEE Transaction on Reliability 2018 68 (1), 317-329. (JCR Q1).
    3. Fan M, **Zeng Z***, Kang R, Zio E and Chen Y. A stochastic hybrid systems model of common-cause failures of degrading components. Reliability Engineering and System Safety 2018; 172: 159-170. (JCR Q1).
    4. Fan M, **Zeng Z***, Kang R and Zio E. Modeling dependent competing failure processes with degradation-shock dependence. Reliability Engineering and System Safety 2017; 165, 422-430. (JCR Q1).

5. Fan M, **Zeng Z***, Kang R, Zio E. and Chen Y. A stochastic hybrid systems based framework for modeling dependent failure processes. PLOS One 2017; 12(2), e0172680. (JCR Q2).
6. Fan M, **Zeng Z**, Kang R and Zio E. Modeling common-cause failures using stochastic hybrid systems. Proceedings of Annual European Safety and Reliability Conference (ESREL2017), Portoroz, Slovenia, 2017.
7. Fan M, **Zeng Z***, Kang R and Zio E. Modeling dependent competing failure processes based on stochastic hybrid systems. Proceedings of Annual European Safety and Reliability Conference (ESREL2016), Glasgow, Scotland, 2016.
8. Fan M, **Zeng Z***, Kang R and Zio E. Reliability modeling of a spool valve considering the dependencies among failure mechanisms and epistemic uncertainty. Proceedings of Annual European Safety and Reliability Conference (ESREL2015), Zurich, Switzerland, 2015.

- Mr. Qingyuan Zhang (Beihang University, China, Sep. 2016 - Dec. 2020):
  - Thesis title: Belief reliability theory and application.
  - Co-supervised (20$\%$) with Prof. Rui Kang (30%) and Dr. Meilin Wen (20%).
  - Graduated with 9 publications in international journals, 4 in international conferences.
  - Recipient of **"Excellence research grand for outstanding PhD students"** from Beihang university.
  - Current position: Postdoc researcher, Beihang University, China.
  - Selected publications with me:
    1. Zhang Q, **Zeng Z***, Zio E, Kang R. Probability box as a tool to model and control the effect of epistemic uncertainty in multiple dependent competing failure processes. Applied Soft Computing. 2017; 56, 570-579. (JCR Q1).
    2. Kang R, Zhang Q, **Zeng Z***, Zio E, Li X. Measuring reliability under epistemic uncertainty: Review on non-probabilistic reliability metrics. Chinese Journal of Aeronautics. 2016; 29(3): 571-579. (JCR Q1).

- Mr. Andrea Belle (Centralesupélec, France, Nov. 2019 - Present):
  - Thesis title: Resilience modelling and optimal protection planning for interconnected railway, electrical and telecommunication systems.
  - Co-supervised (50$\%$) with Prof. Anne Barros (50%).

- Mr. Youba Nait Belaid (Centralesupélec, France, Nov. 2019 - Present):
  - Thesis title: Resilience modelling of interdependent critical infrastructures (CIFRE EDF).

  - Co-supervised (33%) with Prof. Anne Barros (33%) and Dr. Yiping Fang (33%).

- Mr. Rui Li (Centralesupélec, France, Nov. 2020 - Present):
  - Thesis title: Resilience modelling and optimization for 5G infrastructures (CIFRE Orange).
  - Co-supervised (33%) with Prof. Anne Barros (33%) and Dr. Yiping Fang (33%).

- Mr. Khaled Sayad (Centralesupélec, France, Nov. 2020 - Present):
  - Thesis title: Joint optimization of maintenance activities considering interdependency in critical infrastructures (CIFRE Orange).
  - Co-supervised (33%) with Prof. Anne Barros (33%) and Dr. Yiping Fang (33%).

- Mr. Tangfan Xiahou (University of Electronics Science and Technology of China, China, Nov. 2016 - Present):
  - Thesis title: Reliability modeling of complex systems considering epistemic uncertainty.
  - Co-supervised (20$\%$) with Prof. Yu Liu (80%).
  - Selected publications with me:
    1. Xiahou, T., **Zeng Z.**, Liu, Y. Remaining Useful Life Prediction by Fusing Experts' Knowledge and Condition Monitoring Information. IEEE Transactions on Industrial Informatics (Available online). 2020. (JCR Q1).
    2. Xiahou T. (SS) **Zeng Z.**, Liu Y., Huang HZ. Measuring Conflicts of Multi-Source Imprecise Information in Multi-State System Reliability Assessment. IEEE Transactions on Reliability. (JCR Q1, Accepted).

- Miss Yishuang Hu (Zhejiang University, China, Nov. 2018 - Present):
  - Thesis title: Multistate reliability models and optimization methods for electrical systems.
  - Co-supervised (20$\%$) with Prof. Yi Ding (80%).
  - Selected publications with me:
    1. Hu Y., Lin Y., Ding Y., Chen Y., **Zeng Z.** Screening of optimal structure among large-scale multi-state weighted k-out-of-n systems considering reliability evaluation. Reliability Engineering and System Safety. 2020. (JCR Q1).
    2. Ding Y., Hu Y. (SS), Lin Y., **Zeng Z.** Reliability Analysis of Multi-performance Multi-state System Considering Performance-Conversion Process. IEEE Transactions on Reliability. (Accepted).

3.2.2 Master students: 3 completed and 3 on-going

- Andleeb Tariq (May 2020 - Sep. 2020), Supervising master in Nuclear Engineering (100%): Evaluating Accident Propensity of Complex Systems Through Normal Accident Theory and Analytical Hierarchical Process.
- Arpit Shailesh SOLA (May 2020 - Sep. 2020), Supervising master in Nuclear Engineering (100%): Reliability modelling and optimal maintenance planning for steam generator tube failures.
- Jean Meunier-Pion (Sep. 2020 - Sep. 2023), Engineering student from Parcours Recherche (100%): Reputational reliability assessment based on text mining and online customer reviews.
  - Publication: J. Meunier-Pion, J. Liu and Z. Zeng, Big Data Analytics for Reputational Reliability Assessment Using Customer Review Data, Proceeding of ESREL 2021, Anger, France.
- Romain Ray, (Nov. 2020 - March 2021) Master student from Memoire Thematique (100%): Cyber-physical system modelling of smart railway.
- Raghed Saab (Oct. 2020 - Dec. 2020), Internship master student from American University of Benuit, Lebanon (100%): Reliability modeling of railway integration into smart grid. Recipient of **"International internship travel grant"** from French embassy in Lebanon.
- Ameni Ben Amor (Nov. 2020 - March 2021), Master student from Memoire Thematique (100%): Natural language processing and its application in risk and reliability.

3.3 SUPPORTING ACTIVITIES FOR PHD TRAINING

- Co-responsable of Master training (2021 - Present, with Prof. Anne Barros and Dr. Yiping Fang): Risk Resilience and Engineering Management (RREM), Parcours M2 in Master program "Complex Systems Engineering", Centralesupélec, Université Paris-Saclay.
  - ≈ 8 students per year.
  - International master program.
  - ≈ 10 courses, 60 credits.
  - In collaboration with Beihang university (Ecole Centrale Pekin), in terms of double degree agreements (≈ 3 students per year).
- Jury member of research master M2 mention Operation in Master Nulcear Energy (2019 - Present), Centralesupélec, Université Paris-Saclay.
  - ≈ 8 students per year.

  - International master program.
  - In collaboration with INSTN, CEA and Université Paris Sud.

- Co-organizer of PhD school (2015 - 2017): Risk and resilience for complex systems and critical infrastructures, Centralesupélec, Université Paris-Saclay.
  - ≈ 20 students per year.
  - From over 10 countries.
  - Financially supported by T.I.M.E association.

- Invited lecturer, ”Seven weapons you’d better equip yourself before starting an academic journey”, training for first-year PhD students, Beihang University, 2015.
  - Introduce what is research and some fundamental abilities for doing good research.
  - ≈ 20 participants.

3.4 SUMMARY OF TEACHING

Table 3.1 - 3.3 summarizes my teaching activities from 2016 to present. Table 3.1 lists all the courses I taught and their levels. Table 3.2 summarizes my teaching hours for each year. Table 3.3 shows the domains my teachings are involved.

3.5 DETAIL LIST OF TEACHING ACTIVITIES

Sep. 2016 - July 2017:

- Introduction to resilience of complex systems. Elective course for second year engineering students (master level), Centralesupélec, France.
  - ≈ 30 students.
  - Lectures: 12 hrs; Exercise sessions: 3 hrs.

- Co-lecturer for ”Scientific Writing Seminar for PhD Students” at Laboratorie Genie Industriel, Centralesupélec, Feb. 2017.

Sep. 2017 - July 2018:

- Risk analysis. Mandatory course for Master of Nuclear Engineering (MNE), Centralesupélec, France.

Year	Course	Institution	Level
2017	Introduction to resilience of complex systems	Centralesupélec	Master
	Scientific Writing Seminar for PhD Students	Centralesupélec	PhD
2018	Risk analysis	Centralesupélec	Master
	Introduction to risk and reliability	Centralesupélec	Master
	Introduction to resilience of complex systems	Centralesupélec	Master
	Operation management	Centralesupélec	Master
	Maintenance	Centralesupélec	Master
	Risk and resilience for critical infrastructures	Universite Paris-Saclay	PhD
2019	Risk analysis	Centralesupélec	Master
	Introduction to resilience of complex systems	Centralesupélec	Master
	Signal processing for resilience of complex systems	Centralesupélec	Master
	Operation management	Centralesupélec	Master
	Maintenance	Centralesupélec	Master
	Stochastic models	Centralesupélec	Master
	Introduction to reliability engineering	Beihang university	Undergraduate
2020	Risk analysis	Centralesupélec	Master
	Introduction to resilience of complex systems	Centralesupélec	Master
	Operation management	Centralesupélec	Master
	Maintenance	Centralesupélec	Master
	Stochastic models	Centralesupélec	Master
2021	Risk analysis	Centralesupélec	Master
	Signal processing for resilience of complex systems	Centralesupélec	Master
	Maintenance and industry 4.0	Centralesupélec	Master
	Risk identification and control for complex engineering system	Centralesupélec	Master
	Design for resilient system	Centralesupélec	Master

Table 3.1: Summary of my teaching activities (As of July 2021).

Year	Lectures (hours)	Exercises (hours)
2016 - 2017	12	3
2017 - 2018	45	12
2018 - 2019	88	21
2019 - 2020	97	12
2020 - 2021	76	15

Table 3.2: Summary of teaching hours (As of July 2021).

  - ≈ 50 students.
  - Lectures: 3 hrs; Exercise sessions: 3 hrs.

- Introduction to risk and reliability. Mandatory course for first-year engineering students (master level), Centralesupélec, France.
  - ≈ 80 students.
  - Lectures: 9 hrs; Exercise sessions: 6 hrs.

- Introduction to resilience of complex systems. Elective course for second year engineering students (master level), Centralesupélec, France.
  - ≈ 30 students.

Area	Cumulative teaching hours
Risk, reliability and resilience	201
Maintenance optimization	54
Operation management	54
Stochastic modeling	9

Table 3.3: Main areas of teaching (As of July 2021).

  - Lectures: 6 hrs; Exercise sessions: 3 hrs.

- Operation management. Mandatory course for Master of Nuclear Engineering (MNE), Centralesupélec, France.
  - Responsible for the course.
  - ≈ 10 students.
  - Lectures: 18 hrs.
- Maintenance. Mandatory course for Master of Nuclear Engineering (MNE), Centralesupélec, France.
  - ≈ 10 students.
  - Lectures: 9 hrs.
- Co-organizer of PhD school: Risk and resilience for complex systems and critical infrastructures, Centralesupélec, Université Paris-Saclay, with financial support from T.I.M.E. association.

Sep. 2018 - July 2019:

- Risk analysis. Mandatory course for Master of Nuclear Engineering (MNE), Centralesupélec, France.
  - ≈ 50 students.
  - Lectures: 3 hrs; Exercise sessions: 3 hrs.
- Introduction to resilience of complex systems. Elective course for second year engineering students (master level), Centralesupélec, France.
  - ≈ 30 students.
  - Lectures: 6 hrs; Exercise sessions: 3 hrs.
- Signal processing for resilience of complex systems and infrastructures. Mandatory course for first year engineering program (master level). CentraleSupélec, France.
  - ≈ 80 students.
  - Lectures: 6 hrs; Exercise sessions: 6 hrs.
- Challenge week for first year engineering program (master level). CentraleSupélec, France.

  - Supervising student projects.
  - ≈ 20 students.
  - 40 hrs.

- Operation management. Mandatory course for Master of Nuclear Engineering (MNE), Centralesupélec, France.
  - Responsible for the course.
  - ≈ 10 students.
  - Lectures: 18 hrs.

- Maintenance. Mandatory course for Master of Nuclear Engineering (MNE), Centralesupélec, France.
  - Responsible for the course.
  - ≈ 10 students.
  - Lectures: 12 hrs; Exercise sections: 9 hrs.

- Stochastic models. Elective course for second-year engineering students (master level), Centralesupélec, France.
  - ≈ 10 students.
  - Lectures: 3 hrs.

Sep. 2019 - July 2020:

- Risk analysis. Mandatory course for Master of Nuclear Engineering (MNE), Centralesupélec, France.
  - ≈ 50 students.
  - Lectures: 3 hrs; Exercise sessions: 3 hrs.

- Introduction to resilience of complex systems. Elective course for second year engineering students (master level), Centralesupélec, France.
  - ≈ 30 students.
  - Lectures: 15 hrs.

- Challenge week for first year engineering program (master level). CentraleSupélec, France.
  - Supervising student projects.
  - ≈ 20 students.
  - 40 hrs.

- Operation management. Mandatory course for Master of Nuclear Engineering (MNE), Centralesupélec, France.
  - Responsible for the course.
  - ≈ 10 students.
  - Lectures: 18 hrs.
- Maintenance. Mandatory course for Master of Nuclear Engineering (MNE), Centralesupélec, France.
  - Responsible for the course.
  - ≈ 10 students.
  - Lectures: 18 hrs; Exercise sections: 9 hrs.
- Stochastic models. Elective course for second-year engineering students (master level), Centralesupélec, France.
  - ≈ 10 students.
  - Lectures: 3 hrs.

Sep. 2020 - July 2021:

- Risk analysis. Mandatory course for Master of Nuclear Engineering (MNE), Centralesupélec, France.
  - ≈ 50 students.
  - Lectures: 3 hrs; Exercise sessions: 3 hrs.
- Signal processing for resilience of complex systems and infrastructures. Mandatory course for first year engineering program (master level). CentraleSupélec, France.
  - ≈ 80 students.
  - Lectures: 6 hrs; Exercise sessions: 6 hrs.
- Challenge week for first year engineering program (master level). CentraleSupélec, France.
  - Supervising student projects.
  - ≈ 20 students.
  - 40 hrs.
- Maintenance and industry 4.0. Elective course for second-year engineering students (master level), Centralesupélec, France.
  - ≈ 30 students.

  - Lectures: 9 hrs; Exercise sections: 6 hrs.

- Risk identification, mitigation and control for complex engineering system. Elective course for third-year engineering students (master level), Centralesupélec, France.
  - Responsible for the course.
  - ≈ 15 students.
  - Lectures: 15 hrs.

- Design for resilient system. Mandatory course for third-year engineering students (master level), Centralesupélec, France.
  - ≈ 30 students.
  - Exercise sections: 3 hrs.

Chapter 4

SYNTHETIC PRESENTATION OF THE RESEARCH ACTIVITIES

This document contains a synthetic presentation of my research activities to support my application to Habilitation à Diriger des Recherches (HDR) at Université Paris-Saclay. Following the guidelines defined in [1], the document is divided in five sections. Section 1 presents an overview of my main technical and scientific achievements. Section 2 summarizes the recognition from the community I received for my researches. Section 3 presents the international collaborations I developed during my research path. Section 4 lists the research projects/grants I have received/been involved. Section 5 synthetically proposes medium- and long-term plans for my future research. This document is prepared to demonstrate the applicant's compliance to the criteria for obtaining HDR at Université Paris-Saclay, defined in Sect. 1.3 of document [4].

4.1 Overview on my research activities

Modern societies are increasingly relied on large-scale, highly interconnected engineering systems (e.g., power grids, energy distribution network, railway network). Such systems need to be designed with very high degree of reliability and resilience, where reliability refers to the ability of a system to remain operational for a given period of time, under given operation and environmental conditions [153], while resilience means the ability a system to resist and absorb the damages caused by a disruptive events, and quickly recover after the disruption [161]. However, the intrinsic complexity of such complex engineering systems often makes it hard to reach the high requirements on reliability and resilience. My research, then, mainly concerns reliability and resilience of complex engineering systems. More specifically, I have been interested in understanding fundamental failure mechanisms of complex engineering systems and, based on the failure mechanisms, developing new theory, models and techniques to

better quantify and improve reliability and resilience of complex engineering systems. The findings of my researches have been applied on different use cases that represents complex engineering systems from various domains, e.g., nuclear power plants, airplanes, smart grids.

Concerning the general research questions discussed above, my research work can be grouped into five axes. The first axis (Sect. 1.1) is a continuation of my PhD thesis. The main research issue addressed in this axis is to identify the main failure causes of a complex engineering system. Based on knowledge on these failure mechanisms, a new reliability theory, i.e., the belief reliability theory is established. The second - fifth axes (Sect. 1.2 - 1.5) are new research directions I defined after obtaining my PhD degree. Axis 2 (Sect. 1.2) focuses on accurate modeling and efficient analysis of dependent failure behaviors; axis 3 (Sect. 1.3) focuses on modeling epistemic uncertainty and its impact on complex engineering system reliability models; axis 4 (Sect. 1.4) focuses on integrating different data source to make more accurate reliability assessment for a complex engineering system; axis 5 (Sect. 1.5) discusses how to model and improve the resilience of complex engineering systems.

4.1.1 Belief reliability theory and its applications

In risk and reliability, one often uses risk/reliability indexes to support decision making. In traditional risk and reliability approaches, these indexes are mainly estimated from historical data. There are two major drawbacks for these approaches. First, to accurately estimate the risk and reliability indexes, large amount of failure data are required, which is difficult to obtain in practice. Second, purely relied on data makes it difficult to propose design solutions to improve the reliability. In this axis of research, we extend the traditional risk/reliability approaches to jointly consider the influence of design margin, aleatory, and epistemic uncertainty. Our main contribution is the development of a new mathematical theory for integrating design margin, aleatory, and epistemic uncertainty with reliability, called belief reliability theory. Since our first publication on this topic [153], belief reliability has become a new and active research area in reliability theory: Till today, there are over 50 papers, 3 PhD thesis and 1 monograph, published by researchers from all over the world in this area. Below is a list of some representative publications (finished after my PhD) from me and the students I supervised in this area (SS: supervised PhD students, *: Corresponding author):

- **Zeng Z**, Bani-Mustafa T (SS), Flage R, Zio E. An integrated risk index accounting for epistemic uncertainty in Probability Risk Assessment (PRA). Journal of Risk and Reliability 2020. (JCR Q3).
- **Zeng Z**, Kang R, Zio E and Wen M. Uncertainty Theory as a Basis for Belief Reliability. Information Science 2018; 429, 26-36. (JCR Q1).
- **Zeng Z**, Kang R, Zio E and Wen M. A Model-Based Reliability Metric Considering Aleatory and Epistemic Uncertainty. IEEE Access 2017; 5, 15505-15515. (JCR Q1).
- Kang R, Zhang Q (SS), **Zeng Z***, Zio E, Li X. Measuring reliability under epistemic uncertainty: Review on

non-probabilistic reliability metrics. Chinese Journal of Aeronautics. 2016; 29(3): 571-579. (JCR Q1).

4.1.2 Modelling dependent failure behaviors based on stochastic hybrid systems

A fundamental assumption in classic reliability theory is that, failures are independent from one another. This assumption, although greatly simplifies the modelling and analysis, is far from true in reality. How to properly consider dependent failure behaviors, then, becomes a critical issue for the reliability community. The most challenging part when modelling dependent failure behavior is that, the different failure behaviors often involve discrete and continuous variables simultaneously. In this axis of research, this issue is addressed by developing a stochastic hybrid system-based framework for reliability modeling. Failure behavior modelling, efficient simulation and remaining useful life prediction are considered based the developed framework. These results provide a complete toolkit for accurate modelling, efficient analyzing, and better understanding dependent failure behaviors. The researches in axis also involves my PhD student Miss Mengfei Fan. Related publications (* Corresponding author, SS: Supervised student):

- Fan M (SS), **Zeng Z***, Kang R, Zio E and Chen Y. A stochastic hybrid systems model of common-cause failures of degrading components. Reliability Engineering and System Safety 2018; 172: 159-170. (JCR Q1).
- Fan M (SS), **Zeng Z***, Kang R and Zio E. Modeling dependent competing failure processes with degradation-shock dependence. Reliability Engineering and System Safety 2017; 165, 422-430. (JCR Q1).
- **Zeng Z**, Chen Y, Zio, E, Kang R. A compositional method to model dependent failure behaviors based on PoF models. Chinese Journal of Aeronautics. 2017; 30(5): 1729-1739. (JCR Q1).
- Fan M (SS), **Zeng Z***, Kang R, Zio E. and Chen Y. A stochastic hybrid systems based framework for modeling dependent failure processes. PLOS One 2017; 12(2), e0172680. (JCR Q2).
- **Zeng Z**, Kang R, Chen Y. Using PoF models to predict system reliability considering failure collaboration. Chinese Journal of Aeronautics. 2016; 29(5) 1294-1301. (JCR Q1).

4.1.3 Quantifying epistemic uncertainty and its impact on risk and reliability models

In risk and reliability, one often relies models to calculate risk/reliability indexes and support decision making. Epistemic uncertainty, which results from lack of knowledge, exists in the modelling process and affects one's confidence in the model predictions. In classical risk and reliability approaches, epistemic uncertainty is not considered, i.e., the results predicted by the model are fully trusted by the decision maker. In this axis of research, our aim is to develop methods to quantify epistemic uncertainty and integrate it in the models to better support decision-making. This axis also involves my PhD students Ms. Tasneem Bani-Mustafa and Mr. Qingyuan Zhang. Our work established

an integrated framework that allows integrating epistemic uncertainty in risk/reliabilitiy-informed decision-making. Related publications (* Corresponding author, SS: Supervised students):

- Bani-Mustafa T (SS), Flage R, **Zeng Z**, Zio E. An extended method for evaluating assumptions deviations in quantitative risk assessment and application to external flooding risk assessment of a nuclear power plant. Reliability Engineering and Systems Safety 2020; 200, 106947. (JCR Q1).

- Bani-Mustafa T (SS), **Zeng Z**, Zio E, Vasseur D, A practical approach for the evaluation of the strength of knowledge supporting risk assessment models. Safety Science 2020; 124, 104596. (JCR Q1).

- Bani-Mustafa T (SS), **Zeng Z**, Zio E, Vasseur D, A new framework for multi-hazards risk aggregation. Safety Science 2020; 121, 283-302. (JCR Q1).

- Zhang Q (SS), **Zeng Z***, Zio E, Kang R. Probability box as a tool to model and control the effect of epistemic uncertainty in multiple dependent competing failure processes. Applied Soft Computing. 2017; 56, 570-579. (JCR Q1).

- **Zeng Z**, Zio E. A classification-based framework for trustworthiness assessment of quantitative risk analysis. Safety Science. 2017; 99: 215-226. (JCR Q1).

4.1.4 Multi-source data integration for reliability assessment

Traditionally, reliability assessment is based on lifetime data. Collecting enough failure time data, however, is a difficult task in engineering practice, considering the tight constraints on time and resources. On the other hand, as we move into the era of industry 4.0, more and more data are becoming available (e.g., condition-monitoring data from sensors, inspection data, expert judgements, linguistic data from customer reviews). These data, although different in their format and forms of presentations, all contain information regarding product reliability. In this axis of research, I intend to develop unified frameworks that allows integrate data with heterogeneous natures and features for reliability assessment and remaining useful life prediction. Through applications on industrial case studies, the developed models are shown to have provided a new way to make full use of available data and information with different natures for reliability assessments. The researches in this axis also involves my PhD students Miss Jinduo Xing and Mr. Tangfan Xiahou. Related publications (* Corresponding author, SS: Supervised student):

- Xiahou, T. (SS), **Zeng Z.**, Liu, Y. Remaining Useful Life Prediction by Fusing Experts' Knowledge and Condition Monitoring Information. IEEE Transactions on Industrial Informatics (Available online). 2020. (JCR Q1).

- Xing J (SS), **Zeng Z***, Zio E. Dynamic business continuity assessment using condition monitoring data. International Journal of Disaster Risk Reduction 2019, 41, 101334. (JCR Q2).

- Xing J (SS), **Zeng Z***, Zio E. A framework for dynamic risk assessment with condition monitoring data and inspection data. Reliability Engineering and Systems Safety 2019, 191, 106552. (JCR Q1).
- **Zeng Z**, Zio E. Dynamic risk assessment using statistical and condition-monitoring data. IEEE Transactions on Reliability 2018 67 (2), 609-622. (JCR Q1).
- Fan M (SS), **Zeng Z***, Kang R, Zio E and Chen Y. A Sequential Bayesian Approach for Remaining Useful Life Prediction of Dependent Competing Failure Processes. IEEE Transaction on Reliability 2018 68 (1), 317-329. (JCR Q1).
- **Zeng Z**, Di Maio F, Zio E, Kang R. A hierarchical decision making framework for the assessment of the prediction capability of prognostic methods. Journal of Risk and Reliability 2017; 231(1), 36-52. (JCR Q3).

4.1.5 Markov reward models for resilience and business continuity modelling

In resilience and business continuity management, one needs to consider both the potential disruptive events and the financial losses, either directly caused by the disruptive event and/or by the performance losses during the performance disruption period. In this axis, I intend to propose a mathematical framework to support modelling and analyzing processes with such characteristics. Markov reward processes are proposed as a model for resilience and business continuity. Efficient resilience analysis algorithms are also proposed. Related publications (* Corresponding author, SS: Supervised student):

- **Zeng Z**, Fang Y, Zhai Q, Du S. A Markov reward process-based framework for resilience analysis of multi-state energy systems under the threat of extreme events. Reliability Engineering and System Safety. 2021 (Accepted for publication).
- **Zeng Z**, Du S, Ding Y. Resilience Analysis of Multi-state Systems with Time-dependent Behaviors. Applied Mathematical Modeling. 2020; 90, 889-911. (JCR Q1).
- Hu Y. (SS), Lin Y., Ding Y., Chen Y., **Zeng Z.** Screening of optimal structure among large-scale multi-state weighted k-out-of-n systems considering reliability evaluation. Reliability Engineering and System Safety. 2020. (JCR Q1).
- Xing J. (SS), **Zeng Z.***, Zio E., Joint optimization of safety barriers for enhancing business continuity of nuclear power plants against steam generator tube ruptures accidents. Reliability Engineering and Systems Safety. 2020; 202, 107067. (JCR Q1).
- **Zeng Z**, Zio E. An integrated modeling framework for quantitative business continuity assessment. Process Safety and Environmental Protection. 2017; 106: 76-88. (JCR Q1).

- Du S, **Zeng Z***, Cui L, Kang R. Reliability analysis of Markov history-dependent repairable systems with neglected failures. Reliability Engineering and System Safety. 2017; 159: 134-142. (JCR Q1).

4.2 Community recognition

- **Editorial board member**: International Journal of Data Analysis Techniques and Strategies, Oct. 2020 - Present.
  - Scope of journal: Basic/advanced statistics and Bayesian models, data analytics, decision theory, knowledge management, etc.
  - Indexed in Scopus (Elsevier), Academic OneFile (Gale) ACM Digital Library, cnpLINKer (CNPIEC), DBLP Computer Science Bibliography, etc.
  - CiteScore: 1.1 (2019)
  - In charge of selecting reviewers and managing review processes.
- **Leading guest editor**, Applied Science, 2021.
  - JCR indexed, IF: 2.474.
  - Special issue on Modeling dependent failure processes.
  - Co-guest editing with Dr. Jie Liu from Beihang university and Dr. Qingqing Zhai from Shanghai university.
  - 8-10 papers.
- **Qualification for Maî̂tre de Conférences**, CNU Section 61 - Génie informatique, automatique et traitement du signal, Jan. 2019 - Present. (I obtained qualification for MdC 3 years after my PhD defense. This is because I moved to France and did not know about the qualification until 2018.)
- Recipient of **Highly impacted paper award (2016 - 2020)**, Chinese Journal of Aeronautics.
  - For the paper: Kang R, Zhang Q, **Zeng Z***, Zio E, Li X. Measuring reliability under epistemic uncertainty: Review on non-probabilistic reliability metrics. Chinese Journal of Aeronautics (JCR Q1). 2016; 29(3): 571-579. (Citations in Google scholar: 71).
- **Technical program committee member**:
  - Annual European Safety and Reliability Conference (ESREL2014, 2019 - Present).
    - ESREL is one of the most influential conferences in the field of risk and reliability.
    - ≈ 500 participants per year.

    - As a technical program committee member, I am mainly in charge of managing the reviewing process of the submissions, organizing parallel sections, etc.
  - International Conference on System Reliability and Safety (ICSRS, 2018 - Present).
    - ICSRS is an important annual conference in the field of system reliability.
    - ≈ 100 participants per year.
    - As a technical program committee member, I am mainly in charge of managing the reviewing process of the submissions, organizing parallel sections, etc.
  - International Symposium on Reliability Engineering and Risk Management (2021 - Present)
    - ISRERM is an important annual conference in the field of risk and structural reliability.
    - ≈ 100 participants per year.
    - As a technical program committee member, I am mainly in charge of managing the reviewing process of the submissions, organizing parallel sections, etc.

- **Keynote speaker**: 2019 International Forum on Applied Reliability Techniques, Shanghai, China.
  - One of the largest reliability conference focusing on industry problems in China.
  - ≈ 100 participants.
  - Title of my presentation: Reputational reliability assessment based on customer reviews and text mining.
- **Invited panelist**: International Conference on Prognostics and Health Management (ICPHM) 2020.
  - One of the largest conferences in the field of prognostics and reliability.
  - ≈ 300 participants.
  - Title of the panel: Epistemic uncertainty in reliability modeling and optimization.
- **Section chair of international conferences**:
  - Annual European Safety and Reliability Conference (2019 - Present).
    - One of the most influential conferences in the field of risk and reliability.
    - ≈ 500 participants per year.
  - ICPHM2019.
    - ≈ 100 participants.
- **Guest lecturer**, 2019 International Summer School of Aeronautics and Astronautics, Beihang University, China.

  - In charge of a course: Introduction to risk and reliability.
  - 32 hours of teaching.
  - 30 students from different countries.

- **Reviewers of recognized journals:**
  - Reliability Engineering and Systems Safety (≈ 20), top journal in risk and reliability.
  - IEEE Transactions on Reliability (≈ 10), top journal in risk and reliability.
  - Quanlity and Reliability Engineering International (≈ 5), highly impacted journal in risk and reliability.
  - Journal of Risk and Reliability (≈ 5), highly impacted journal in risk and reliability.
  - Maintenance and reliability (≈ 5), highly impacted journal in risk and reliability.
  - Applied Mathematical Modeling (≈ 5), top journal in applied mathematics and operations research.
  - IISE Transactions (≈ 5), top journal in industrial engineering.
  - Computers & Industrial Engineering (≈ 5), top journal in industrial engineering.
  - IEEE Transactions on Industrial Electronics (≈ 5), top journal in electrical engineering and computer science.
  - IEEE Transactions on Industrial Informatics (≈ 5), top journal in electrical engineering and computer science.
  - and many others.

- **Invited seminars:**
  - Prognostics and health management by integrating condition-monitoring data and expert judgment, UQSay - An open seminar jointly orgnaized by Laboratoires L2S and MSS-Mat, Centralesupélec, France, 2019. ≈ 30 participants.
  - Reliability assessment through text mining.
    * Beihang university, China. 2018. ≈ 30 participants.
    * Forum of high-end oversea experts. Shanghai university, China. 2018. ≈ 10 participants.
  - Modeling dependent failure behaviors through stochastic hybrid systems.
    * Forum of management science, Fuzhou university, China. 2018. ≈ 20 participants.
    * University of Electronic Techonology of China, China. 2018. ≈ 20 participants.
    * Zhejiang university, China. 2018. ≈ 20 participants.
    * Sichuan university, China. 2018. ≈ 20 participants.
  - and many others.

4.3 International collaborations

New collaborations after my PhD defense:

- Prof. David Coit, Rutgers University, United States:
  - To host Prof. Coit to visit Centralesupélec for one month (2021, receive financial support from Centralesupélec).
- Prof. Yi Ding, Zhejiang University, Zhejiang, China:
  - Invited by Prof. Ding to visit Zhejiang University, China for one weeks (July 2019).
  - Co-supervising students: Ms. Yishuang Hu.
  - Co-authoring papers:
    1. **Zeng Z**, Du S, Ding Y. Resilience Analysis of Multi-state Systems with Time-dependent Behaviors. Applied Mathematical Modeling. 2020; 90, 889-911. (JCR Q1).
    2. Hu Y., Lin Y., Ding Y., Chen Y., **Zeng Z.** Screening of optimal structure among large-scale multi-state weighted k-out-of-n systems considering reliability evaluation. Reliability Engineering and System Safety. 2020. (JCR Q1).
- Prof. Roger Flage, University of Stavanger, Norway:
  - Host Prof. Flage to visit Centralesupélec for one month (2018, received financial support from Centralesupélec).
  - Co-authoring papers:
    1. **Zeng Z**, Bani-Mustafa T, Flage R, Zio E. An integrated risk index accounting for epistemic uncertainty in Probability Risk Assessment (PRA). Journal of Risk and Reliability 2020. (JCR Q3).
    2. Bani-Mustafa T, Flage R, **Zeng Z**, Zio E. An extended method for evaluating assumptions deviations in quantitative risk assessment and application to external flooding risk assessment of a nuclear power plant. Reliability Engineering and Systems Safety 2020; 200, 106947. (JCR Q1).
- Prof. Yu Liu, University of Electronics Science and Technology of China, Sichuan, China:
  - Invited by Prof. Yu Liu to visit University of Electronics Science and Technology of China, Sichuan, China for two weeks (July 2019).
  - Co-supervising students: Mr. Tangfan Xiahou.
  - Co-authoring papers:

1. Xiahou, T., **Zeng Z.**, Liu, Y. Remaining Useful Life Prediction by Fusing Experts' Knowledge and Condition Monitoring Information. IEEE Transactions on Industrial Informatics (Available online). 2020. (JCR Q1).

- Prof. Qingqing Zhai, Shanghai University, Shanghai, China:
  - Invited by Prof. Qingqing Zhai to visit Shanghai University, Shanghai, China for one month (August 2019).
  - Co-authoring papers:
    1. **Zeng Z**, Fang Y, Zhai Q, Du S. A Markov reward process-based framework for resilience analysis of multistate energy systems under the threat of extreme events. Reliability Engineering and System Safety. 2021 (Accepted for publication).
    2. Shijia Du, **Zhiguo Zeng**, Yiping Fang, Qingqing Zhai. Resilience analysis of multistate systems based on Markov reward processes. ICSRS 2019. Rome, Italy, 2019.

Collaborations as a continuation of my PhD thesis:

- Prof. Rui Kang, Beihang University, Beijing, China:
  - Co-supervising students:
    * Ms. Mengfei Fan
    * Mr. Qingyuan Zhang
  - Co-authoring papers:
    1. Zio E, Fan M, **Zeng Z***, Kang R, Application of reliability technologies in civil aviation: lessons learnt and perspectives. Chinese Journal of Aeronautics. 2019 (32) 1: 143-158. (JCR Q1).
    2. Fan M, **Zeng Z***, Kang R, Zio E and Chen Y. A Sequential Bayesian Approach for Remaining Useful Life Prediction of Dependent Competing Failure Processes. IEEE Transaction on Reliability 2018 68 (1), 317-329. (JCR Q1).
    3. Fan M, **Zeng Z***, Kang R, Zio E and Chen Y. A stochastic hybrid systems model of common-cause failures of degrading components. Reliability Engineering and System Safety 2018; 172: 159-170. (JCR Q1).
    4. Fan M, **Zeng Z***, Kang R and Zio E. Modeling dependent competing failure processes with degradation-shock dependence. Reliability Engineering and System Safety 2017; 165, 422-430. (JCR Q1).
    5. Fan M, **Zeng Z***, Kang R, Zio E. and Chen Y. A stochastic hybrid systems based framework for modeling dependent failure processes. PLOS One 2017; 12(2), e0172680. (JCR Q2).

6. Fan M, **Zeng Z**, Kang R and Zio E. Modeling common-cause failures using stochastic hybrid systems. Proceedings of Annual European Safety and Reliability Conference (ESREL2017), Portoroz, Slovenia, 2017.
7. Fan M, **Zeng Z***, Kang R and Zio E. Modeling dependent competing failure processes based on stochastic hybrid systems. Proceedings of Annual European Safety and Reliability Conference (ESREL2016), Glasgow, Scotland, 2016.
8. Fan M, **Zeng Z***, Kang R and Zio E. Reliability modeling of a spool valve considering the dependencies among failure mechanisms and epistemic uncertainty. Proceedings of Annual European Safety and Reliability Conference (ESREL2015), Zurich, Switzerland, 2015.
9. Zhang Q, **Zeng Z***, Zio E, Kang R. Probability box as a tool to model and control the effect of epistemic uncertainty in multiple dependent competing failure processes. Applied Soft Computing. 2017; 56, 570-579. (JCR Q1).
10. Kang R, Zhang Q, **Zeng Z***, Zio E, Li X. Measuring reliability under epistemic uncertainty: Review on non-probabilistic reliability metrics. Chinese Journal of Aeronautics. 2016; 29(3): 571-579. (JCR Q1).

4.4 Projects and grants

Research projects independent from my PhD thesis:

- Collaborative research project with GE medical care: **Participant,** 2021 - Present.
  - Title: Reliability modeling and maintenance optimization for recycled critical components based on incomplete data.
  - In this project, I am in charge of:
    * developing reliability models and maintenance optimization models;
    * designing a project for education purposes for engineering students in Centralesupélec, based on my course Mainenance and Industry 4.0.
- Visiting grants for outstanding scholars, Shanghai University, €7,500, **Grant holder**, July 2019 - August 2019.
  - Title: Reliability modeling and maintenance optimization considering dependent failure behaviors.
  - Invited to work in Shanghai University for one month.
  - Main objectives of this project:
    * develop component reliability models considering complex multiple dependent failure processes;

    - optimize condition-based maintenance plans for systems with dependent failure behaviors;
    - give academic seminars and supervise students together.

- Natural Science Foundation of China, Starting grant, €25,000, **Participant**, 2017 - 2019.
  - Title: Resilience modeling based on aggregated Markov process.
  - In this project, I am in charge of:
    - collaborate with the grant holder to develop Markov reward process models for resilience;
    - develop efficient simulation algorithms for resilience analysis based on semi-Markov reward models;
    - find real-world applications to validate the theoretical frameworks.

Teaching project independent from my PhD thesis:

- European project Erasmus+ Key Action 2, **Participant**, 2019 - 2022.
  - Title: INTegrated SYStem for European Digital learning.
  - Leaded by Techical University of Berlin, this projects aim at developing a handbook and an platform to support online-learning of industrial engineering in the European level.
  - As one of the representatives of Centralesupélec, I am in charge of:
    - validate the handbook by following it to design and implement an e-learning course: Risk Identification, Mitigation and Control for complex Engineering System (Elective course for 3rd year engineering students at Centralesupélec, March 2021).
    - develop a serious game to support the e-learning:
      - designed as a mystery murder game;
      - the students play as system designers;
      - some potential failures are injected into a system, based on some real-world accidents;
      - the students are asked to identify them through the serious game.

Research projects related to my PhD thesis:

- Natural Science Foundation of China, Career grant, €160,000, **Participant**, 2016 - 2019.
  - Title: Belief reliability metrics and related analysis methods.
  - In this project, I am in charge of:
    - write the project proposal;
    - define belief reliability metrics based on uncertainty theory;
    - develop belief reliability analysis metrics based on fault tree and unvertainty theory.

Chapter 5

COMPLETE AND CLASSIFIED LIST OF PUBLICATIONS AND COMMUNICATIONS

5.1 RESEARCH OUTPUTS: QUANTITATIVE SUMMARY

Type	Related to PhD thesis	Independent from PhD thesis	Total
Journal papers (JCR indexed)	4	28	32
International Conference papers	9	14	23
Book/Book chapters	1	2	3

Table 5.1: Quantitative summary of scientific outputs (Data of 12/03/2021).

	Google scholar	Web of Science	Scopus
Citations	900	640	759
H-index	17	14	16

Table 5.2: Citations and H-indexes (Data of 12/07/2022).

5.2 COMPLETE LIST OF PUBLICATIONS

This section presents a complete list of my publications. In the list, * indicates corresponding author, and SS represents supervised PhD students.

5.2.1 Journal Papers

Online reliability assessment and remaining useful life prediction for industry 4.0

1. Xiahou, T. (SS), **Zeng Z.**, Liu, Y. Remaining Useful Life Prediction by Fusing Experts' Knowledge and Condition Monitoring Information. IEEE Transactions on Industrial Informatics (Available online). 2020. (JCR Q1).

2. Xing J (SS), **Zeng Z***, Zio E. Dynamic business continuity assessment using condition monitoring data. International Journal of Disaster Risk Reduction 2019, 41, 101334. (JCR Q2).

3. Xing J (SS), **Zeng Z***, Zio E. A framework for dynamic risk assessment with condition monitoring data and inspection data. Reliability Engineering and Systems Safety 2019, 191, 106552. (JCR Q1).

4. Zio E, Fan M (SS), **Zeng Z***, Kang R, Application of reliability technologies in civil aviation: lessons learnt and perspectives. Chinese Journal of Aeronautics. 2019 (32) 1: 143-158. (JCR Q1).

5. **Zeng Z**, Zio E. Dynamic risk assessment using statistical and condition-monitoring data. IEEE Transactions on Reliability 2018 67 (2), 609-622. (JCR Q1).

6. Fan M (SS), **Zeng Z***, Kang R, Zio E and Chen Y. A Sequential Bayesian Approach for Remaining Useful Life Prediction of Dependent Competing Failure Processes. IEEE Transaction on Reliability 2018 68 (1), 317-329. (JCR Q1).

7. **Zeng Z**, Di Maio F, Zio E, Kang R. A hierarchical decision making framework for the assessment of the prediction capability of prognostic methods. Journal of Risk and Reliability 2017; 231(1), 36-52. (JCR Q3).

Dependent competing failure process: Modeling and analysis

8. Fan M (SS), **Zeng Z***, Kang R, Zio E and Chen Y. A stochastic hybrid systems model of common-cause failures of degrading components. Reliability Engineering and System Safety 2018; 172: 159-170. (JCR Q1).

9. Chen J, Zio E, Li J, **Zeng Z***, Chong Bu. Accelerated life test for reliability evaluation of pneumatic cylinders. IEEE Access 2018; 6, 75062-75075. (JCR Q1).

10. Fan M (SS), **Zeng Z***, Kang R and Zio E. Modeling dependent competing failure processes with degradation-shock dependence. Reliability Engineering and System Safety 2017; 165, 422-430. (JCR Q1).

11. **Zeng Z**, Chen Y, Zio, E, Kang R. A compositional method to model dependent failure behaviors based on PoF models. Chinese Journal of Aeronautics. 2017; 30(5): 1729-1739. (JCR Q1).

12. Fan M (SS), **Zeng Z***, Kang R, Zio E. and Chen Y. A stochastic hybrid systems based framework for modeling dependent failure processes. PLOS One 2017; 12(2), e0172680. (JCR Q2).

13. **Zeng Z**, Kang R, Chen Y. Using PoF models to predict system reliability considering failure collaboration. Chinese Journal of Aeronautics. 2016; 29(5) 1294-1301. (JCR Q1).

14. Chen Y, **Zeng Z***, Kang R. Validation methodology for distribution-based degradation model. Journal of Systems Engineering and Electronics, 2012; 23(4): 553-559. (JCR Q4).

Uncertainty modeling and analysis in reliability engineering

15. **Zeng Z**, Bani-Mustafa T (SS), Flage R, Zio E. An integrated risk index accounting for epistemic uncertainty in Probability Risk Assessment (PRA). Journal of Risk and Reliability 2020. (JCR Q3).

16. Bani-Mustafa T (SS), Flage R, **Zeng Z**, Zio E. An extended method for evaluating assumptions deviations in quantitative risk assessment and application to external flooding risk assessment of a nuclear power plant. Reliability Engineering and Systems Safety 2020; 200, 106947. (JCR Q1).

17. Bani-Mustafa T (SS), **Zeng Z**, Zio E, Vasseur D, A practical approach for the evaluation of the strength of knowledge supporting risk assessment models. Safety Science 2020; 124, 104596. (JCR Q1).

18. Bani-Mustafa T (SS), **Zeng Z**, Zio E, Vasseur D, A new framework for multi-hazards risk aggregation. Safety Science 2020; 121, 283-302. (JCR Q1).

19. **Zeng Z**, Kang R, Zio E and Wen M. Uncertainty Theory as a Basis for Belief Reliability. Information Science 2018; 429, 26-36. (JCR Q1).

20. **Zeng Z**, Kang R, Zio E and Wen M. A Model-Based Reliability Metric Considering Aleatory and Epistemic Uncertainty. IEEE Access 2017; 5, 15505-15515. (JCR Q1).

21. Zhang Q (SS), **Zeng Z***, Zio E, Kang R. Probability box as a tool to model and control the effect of epistemic uncertainty in multiple dependent competing failure processes. Applied Soft Computing. 2017; 56, 570-579. (JCR Q1).

22. **Zeng Z**, Zio E. A classification-based framework for trustworthiness assessment of quantitative risk analysis. Safety Science. 2017; 99: 215-226. (JCR Q1).

23. Guo M, Fan M (SS), **Zeng Z**, Wen M and Kang R. Evaluation on the Effect of Reliability Simulation Tests Based on Experts? Information Fusion Method (in Chinese). Electronic Science and Technology, 2016; 3(2): 314-320.

24. Kang R, Zhang Q (SS), **Zeng Z***, Zio E, Li X. Measuring reliability under epistemic uncertainty: Review on non-probabilistic reliability metrics. Chinese Journal of Aeronautics. 2016; 29(3): 571-579. (JCR Q1).

25. Jiang X, **Zeng Z***, Kang R and Chen Y. A Naive Bayes Based Method for Evaluation of Electronic Product Reliability Simulation Tests (in Chinese). Electronic Science and Technology, 2015; 2(1): 49-54.

26. Fan M (SS), **Zeng Z*** and Kang R. A novel approach to measure reliability based on belief reliability (in Chinese). Journal of Systems Engineering and Electronics, 2015, 37(11):2648-2653.

27. **Zeng Z**, Wen M*, Kang R. Belief reliability: A new metrics for products' reliability. Fuzzy Optimization and Decision Making 2013; 12(1): 15-27. (JCR Q1).

Advanced models for repairable/resilient systems

28. **Zeng Z**, Fang Y, Zhai Q, Du S. A Markov reward process-based framework for resilience analysis of multistate energy systems under the threat of extreme events. Reliability Engineering and System Safety (JCR Q1). 2021 (Accepted for publication).

29. **Zeng Z**, Du S, Ding Y. Resilience Analysis of Multi-state Systems with Time-dependent Behaviors. Applied Mathematical Modeling. 2020; 90, 889-911. (JCR Q1).

30. Hu Y. (SS), Lin Y., Ding Y., Chen Y., **Zeng Z.** Screening of optimal structure among large-scale multi-state weighted k-out-of-n systems considering reliability evaluation. Reliability Engineering and System Safety. 2020. (JCR Q1).

31. Xing J. (SS), **Zeng Z.***, Zio E., Joint optimization of safety barriers for enhancing business continuity of nuclear power plants against steam generator tube ruptures accidents. Reliability Engineering and Systems Safety. 2020; 202, 107067. (JCR Q1).

32. **Zeng Z**, Zio E. An integrated modeling framework for quantitative business continuity assessment. Process Safety and Environmental Protection. 2017; 106: 76-88. (JCR Q1).

33. Du S, **Zeng Z***, Cui L, Kang R. Reliability analysis of Markov history-dependent repairable systems with neglected failures. Reliability Engineering and System Safety. 2017; 159: 134-142. (JCR Q1).

34. Ding Y., Hu Y. (SS), Lin Y., **Zeng Z.** Reliability Analysis of Multi-performance Multi-state System Considering Performance-Conversion Process. IEEE Transactions on Reliability. (JCR Q2, Accepted).

35. Xiahou T. (SS) **Zeng Z.**, Liu Y., Huang HZ. Measuring Conflicts of Multi-Source Imprecise Information in Multi-State System Reliability Assessment. IEEE Transactions on Reliability. (JCR Q1, Accepted).

Under Review / Revision

1. Gao J., Liu Y., Xiahou T. and **Zeng Z.**, Joint Optimization of Inspection and Selective Maintenanace of Multi-state System through Deep Reinforcement Learning. IISE Transactions. (Under review).

2. Wu M., Xiahou T. (SS), Liu Y., **Zeng Z.** Remaining Useful Life Prediction under Imprecise Observations: An Interval Particle Filtering Approach. IISE Transactions. (Under 1st round of review).

3. Hu Y. (SS), Ding Yi., **Zeng Z.** Redundancy optimization for multi-state series-parallel systems using ordinal optimization-based-genetic algorithm. Journal of Risk and Reliability. (Under 1st round of review).

5.2.2 International Conference Papers

1. Shijia Du, **Zhiguo Zeng**. Resilience analysis of multistate energy system with time-dependent behaviors. Proceedings of the 30th European Safety and Reliability Conference (ESREL). Vinece, Italy, 2020.

2. Shijia Du, **Zhiguo Zeng**, Yiping Fang, Qingqing Zhai. Resilience analysis of multistate systems based on Markov reward processes. ICSRS 2019. Rome, Italy, 2019.

3. **Zeng Z**, Zio E., Assessing reliability reputation of products based on online customer reviews. Proceedings of the 29th European Safety and Reliability Conference (ESREL). Hanover, Germany, 2019.

4. Zeng Z, Zio E. Joint optimization of business continuity by designing safety barriers for accident prevention, mitigation and emergency responses, ICSRS2018, Bacelona, Spain, 2018.

5. Zeng Z, Zio E, Modelling Unexpected Failures with a Hierarchical Bayesian Model, ICSRS2017, Milano, 2017.

6. Bani-Mustafa T, Zeng Z, Zio E and Vasseur D, A Framework for Multi-Hazards Risk Aggregation Considering Risk Model Maturity Levels, ICSRS2017, Milano, 2017.

7. Bani-Mustafa T, **Zeng Z**, Zio E, Vasseur D, Strength of Knowledge Assessment for Risk Informed Decision Making. Proceedings of Annual European Safety and Reliability Conference (ESREL2018), Trondheim, Norway, 2018.

8. Du S, **Zeng Z**, Kang R and Zio E. A multistate model for resilience analysis of a distributed generation system. Proceedings of the 10th international conference on mathematical methods in reliability, Grenoble, France, 2017.

9. **Zeng Z** and Zio E. Interval-valued importance measures for business continuity management. Proceedings of Annual European Safety and Reliability Conference (ESREL2017), Portoroz, Slovenia, 2017.

10. Du S, **Zeng Z**, Kang R and Zio E. Time-dependent reliability assessment of a distributed generation system based on multi-valued decision diagrams and Markov processes. Proceedings of Annual European Safety and Reliability Conference (ESREL2017), Portoroz, Slovenia, 2017.

11. Xing J, **Zeng Z** and Zio E. An integrated framework for condition-informed probabilistic risk assessment. Proceedings of Annual European Safety and Reliability Conference (ESREL2017), Portoroz, Slovenia, 2017.

12. Fan M, **Zeng Z**, Kang R and Zio E. Modeling common-cause failures using stochastic hybrid systems. Proceedings of Annual European Safety and Reliability Conference (ESREL2017), Portoroz, Slovenia, 2017.

13. Du S, **Zeng Z***, Kang R and Zio E. Resilience modeling of multi-state systems based on aggregated stochastic processes. Proceedings of Annual European Safety and Reliability Conference (ESREL2016), Glasgow, Scotland, 2016.

14. Fan M, **Zeng Z***, Kang R and Zio E. Modeling dependent competing failure processes based on stochastic hybrid systems. Proceedings of Annual European Safety and Reliability Conference (ESREL2016), Glasgow, Scotland, 2016.

15. Fan M, **Zeng Z***, Kang R and Zio E. Reliability modeling of a spool valve considering the dependencies among failure mechanisms and epistemic uncertainty. Proceedings of Annual European Safety and Reliability Conference (ESREL2015), Zurich, Switzerland, 2015.

16. **Zeng Z***, Kang R, Wen M, et al. Measuring reliability during product development considering aleatory and epistemic uncertainty. Proceedings of the 61st Annual Reliability and Maintainability Symposium, Palm Harbor, 2015.

17. **Zeng Z***, Zhang Q, Kang R. Reliability Box as a Tool for Reliability Analysis in Presence of Epistemic Uncertainty. Proceeding of the Ninth International Conference on Mathematical Methods in Reliability (MMR2015), Tokyo, 2015.

18. **Zeng Z***, Kang R, Chen Y. Life oriented design (LOD): Using time-to-failure distribution as the objective of quantitative reliability design. Proceeding of the 8th International Conference on Modelling in Industrial Maintenance and Reliability (MIMAR) Oxford, UK, 2014.

19. **Zeng Z***, Kang R, Chen Y. A physics-of-failure-based approach for failure behavior modeling: With a focus on failure collaborations. Proceedings of Annual European Safety and Reliability Conference (ESREL2014), Wroclaw, Poland, 2014.

20. **Zeng Z***, Chen Y, Kang R. Failure behavior modeling: Towards a better characterization of product failures. Proceedings of 2013 IEEE International Conference on Prognostics and Health Management, Milano, Italy. Chemical Engineering Transactions 2013; 33: 571-576.

21. **Zeng Z***, Chen Y, Kang R. Simulation-based constructions of reliability confidence intervals from degradation data. Proceedings of 2013 International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering (QR2MSE), pp. 908-913. IEEE, 2013.

22. **Zeng Z***, Chen Y, Kang R. The effects of material degradation on the sealing performances of O-rings. Proceedings of teh 3rd International Conference on Mechanical Science and Engineering, Hong Kong, China, 2013.

23. **Zeng Z***. Failure rates of some typical belief reliability functions. Proceedings of International Conference on Information Science and Management, Dunhuang, China, 2012.

5.2.3 Book / Book chapters

- Mengfei Fan and Zhiguo Zeng, Reliabilty modeling, analysis and dynamic assessment considering dependent failure behaviors, National Defense Industry Publishing House, Beijing China, 2021 (to appear), in Chinese.
- Book chapters: Reliability techniques for aviation products with high reliability and long life (in Chinese), 2 chapters, 2020.
- Book chapters: Belief reliability theory and methods (in Chinese), 2 chapters, 2020.

Part II

PAST RESEARCH ACTIVITIES AND FUTURE PLANS

Chapter 6

SEEKING CERTAINTY OUT OF UNCERTAINTY: A NEW FRAMEWORK FOR MEASURING RELIABILITY

This chapter summarizes some of my representative research results in research axis 1. The focus of this axis is to develop a conceptual framework for understanding failure causes and, further, use it to quantify risk and reliability. In Sect. 6.1, we briefly review the related literature and define the research questions considered in this axis. Sections 6.2-6.4 present some representative results in this axis. In Sect. 6.2, a conceptual framework is developed to explain failure causes. New risk and reliability metrics are developed in Sects. 6.3 and 6.4 based on the developed conceptual framework. Finally, in Sect. 6.5, we summarize the major contributions achieved in this research axis.

6.1 Research questions

Measuring reliability refers to quantifying the reliability of a component or system by quantitative metrics. Unlike measuring physical quantities like electrical current, a challenging issue in measuring reliability is that, usually, reliability is associated with large amount of uncertainty: different samples never fail at exactly the same time. How to deal with uncertainty, is, then, a critical issue in reliability measurement. On the other hand, there is also certainty out of uncertainty: if one examine each failure carefully, usually we could identify deterministic failure causes. How to integrate the certain knowledge on failure causes with the uncertainty, is, then, a even more challenging problem.

In the early years of reliability engineering development, reliability has been measured by probability-based metrics, e.g., in terms of the probability that the component or system does not fail (referred to as probabilistic reliability [21]), and estimated by statistical methods based on failure data (e.g., see [98]). However, in engineering

practice, the available failure data, if there are any, are often far from sufficient for accurate statistical estimates [14]. Also, the statistical methods do not explicitly model the actual process that leads to the failure. Rather, the failure process is regarded as a black box and assumed to be uncertain, described indirectly based on the observed distribution of the Time-To-Failure (TTF). From the perspective of uncertainties, broadly speaking, uncertainty can be categorized as aleatory uncertainty, which refers to the uncertainty inherent in the physical behavior of the system and epistemic uncertainty, which refers to the uncertainty that is caused by incomplete knowledge [34]. The statistical methods do not separate the root causes of failures and uncertainties and therefore, they do not distinguish between aleatory and epistemic uncertainties.

As technology evolves, modern products often have high reliability, making it even harder to collect enough failure data, which severely challenges the use of statistical methods [33]. At the same time, as the knowledge of the failure mechanisms accumulates, deterministic models are available to describe the failure process based on the physical knowledge of the failure mechanisms (referred to as physics-of-failure (PoF) models [97]). An alternative method to estimate the probabilistic reliability is, then, that based on the PoF models. In this paper, these methods are referred to as the model-based methods. Unlike statistical methods, model-based methods treat the actual failure process as a white box: the TTFs are predicted by deterministic PoF models, while the uncertainty affecting the TTF is assumed to be caused by random variations in the model parameters (aleatory uncertainty). The probabilistic reliability is, then, estimated by propagating aleatory uncertainties through the model analytically or numerically, e.g., by Monte Carlo simulation [101]. Compared to statistical methods, model-based methods explicitly describe the actual failure process (by the deterministic PoF models) and separate the root cause of failures (assumed to be deterministic) and the aleatory uncertainty (the random variation of model parameters). The separation of deterministic root causes and aleatory uncertainty allows the designer to implement parametric design for reliability, e.g., the Reliability-Based Design Optimization (RBDO) [113], tolerance optimization [163], etc., which marks a significant advancement in reliability engineering.

From the perspective of uncertainties, only aleatory uncertainty is considered in the model-based methods. In practice, however, the trustfulness of the predicted reliability is severely influenced by epistemic uncertainty. As in today's highly competitive markets, it is more and more frequent to use the model-based method to measure reliability, due to the severe shortage on failure data. To better quantify the reliability with the model-based methods, the effect of epistemic uncertainty should also be considered. Epistemic uncertainty relates to the completeness and accuracy of the knowledge: if the failure process is poorly understood, there will be large epistemic uncertainty [11]. For instance, the deterministic PoF model might not be able to perfectly describe the failure process, e.g., due to incomplete understanding of the failure causes and mechanisms [11]. Besides, the precise values of the model parameters might not be accurately estimated due to lack of data in the actual operational and environmental conditions. Both of these two factors introduce epistemic uncertainty into the reliability estimation: the more severe the effect of these factors is, the less trustful the predicted reliability is.

In literature, there are various approaches to measure reliability under epistemic uncertainty, e.g., probability theory (subjective interpretation [65]), evidence theory [15], interval analysis [143], fuzzy interval analysis [87], possibility theory [75], etc. Two issues, however, still remain to be addressed:

1. Most of the existing researches on epistemic uncertainty focus on technical aspects of propagating the uncertainty in risk and reliability models. However, before addressing the technical problems, a conceptual framework that well explains failure causes and includes the certain and uncertain contributors, is needed.
2. Existing works often consider the epistemic uncertainty and its impact separately. New risk/reliability metrics are needed to integrate the contributions of deterministic failure causes, aleatory and epistemic uncertainty.

In this chapter, we focus on these two research questions. Section 6.2 addresses the first question by proposing a generic, conceptual model for failure causes. Sections 6.3 and 6.4 address the second research question, by proposing a new metric for reliability and risk, respectively.

6.2 A generic conceptual framework for failure causes

In this section, we develop a generic, conceptual framework in Figure 6.1 to understand major causes to failures and the relationship among them. Although different definitions of failure can be found in literature, most of them were coalesced around the terminology published in 1990 by the International Electrotechnical Commission (IEC) and subsequently adopted by a number of international standards:

> "failure is defined as the termination of the ability of an item to perform a required function [64]."

Functional thresholds are often defined, for each performance parameter, such that when the performance parameter lies within the associated functional threshold, the system is in normal functioning state. Here, we define performance margin as the difference between the value of the performance parameter and the associated functional threshold. Then, failure can be represented by the performance margin. By stating this, we are making the following arguments: for any failure, one can always define one or several performance margins, where failure occurs whenever the performance margin(s) is less than zero. For example, performance margin of a mechanical structure can be defined as the difference between the mechanical stress and material strength (in mechanical engineering, performance margin is often referred to as safety margin) [97]. In [86], the performance margin of a lock mechanism is defined based on a kinetic modeling as the difference between the angle error and its allowable limit.

Based on whether the failure is known and considered in the design phase of the system, a failure can be classified as conscious failure and blind failure. Designers usually have good understanding about the conscious failure, its causes and the associated performance margin, so that the failure can be considered (explicitly or implicitly) in the reliability design. Broadly speaking, the causes of an conscious failure include design margin, degradation and aleatory uncertainty of the performance margin:

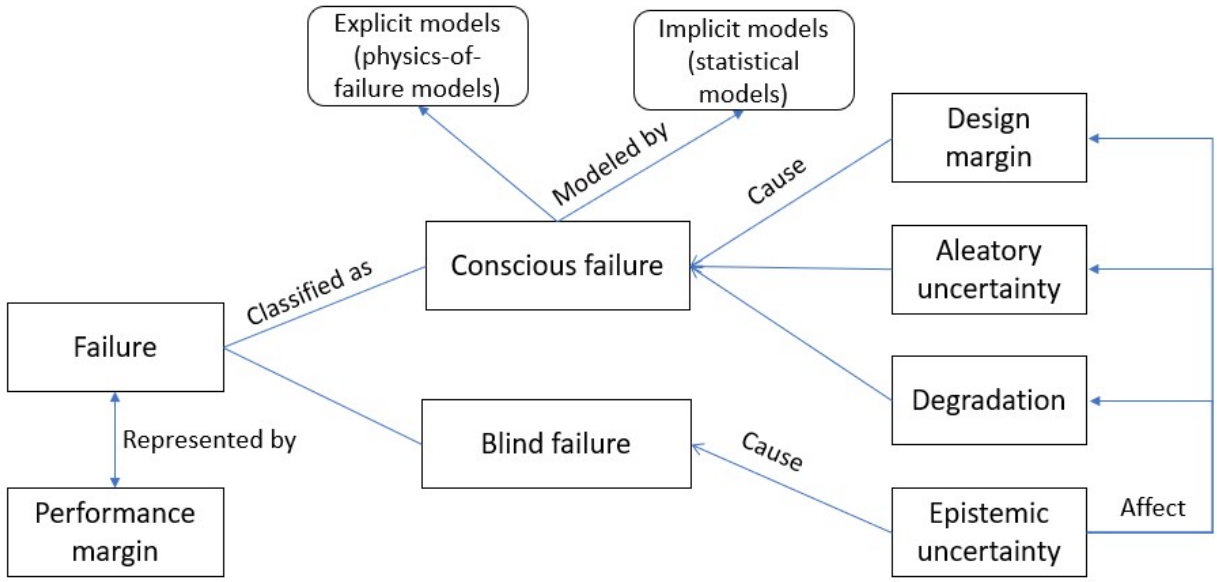


Figure 6.1: A conceptual framework for failure causes.

- Design margin is the nominal values of the performance margin that have to be chosen by the designer. By choosing a larger design margin, the designer can have more confidence on the reliability, while paying the price of potential over-design. For example, safety factors are often used in mechanical design to guide the choice of material strength, so that a high degree of design margin can be guaranteed [31].

- Aleatory uncertainty refers to the inherent randomness in product operation and failure processes [34]. Aleatory uncertainty might arise from various sources, *e.g.*, capability of the production process, accuracy of the machine, variations in the source materials [31]. As a result, the design margin might deviate from its designed value, creating potential rooms for failures. Such a failure cause has been studies extensively as stress-strength interference [62].

- By carefully tuning the design margin and aleatory uncertainty, one is able to ensure the reliability at $t = 0$. However, various performance degradation mechanisms might cause the performance margin to degrade over t. Hence, a reliable product at $t = 0$ might gradually becomes unreliable when t grows. Examples of degradation mechanisms include wear, corrosion, erosion, creep, fatigue, *etc* [42].

By considering the design margin, degradation, and aleatory uncertainty, either explicitly or implicitly, models can be developed to quantify reliability and support reliability-based design. Explicit models derive the performance margin and its evolution over time based on physics of failure [156]. Reliability of the product can, then, be calculated by considering the uncertainty in the physics-based performance margin model [49]. Implicit models, on the other hand, directly estimate the reliability based on the historical time-to-failure data, without explicitly modeling the performance margin [98]. Both the explicit and implicit model share an important underlying assumption, although

rarely discussed explicitly: the predicted reliability only covers the conscious failures. This fact creates an interesting paradox for the current reliability modeling and estimation: to model and estimate the reliability, we need to know what failures are going to happen (conscious failures). But if a designer already knows a failure is going to happen, usually he/she should have already taken measures to prevent it, or, at least, control its consequence. If so, why we still have failures occurring in practice?

In fact, a large number of failures that occur in practice belong to blind failures, *i.e.,* the failures that are unknown or not considered in the design phase. For example, In the 1950s, two consecutive mid-air explosions of the airplane De Havilland Comet occurred, which were found out to be caused by metal fatigue. However, before these accidents, people did not understand that metal fatigue could occur in such a manner [137]. As shown in Figure 6.1, blind failure is mainly caused by epistemic uncertainty, *i.e.,* the uncertainty that arises from lack of knowledge [34]. In reliability engineering, testing like Highly Accelerated Life Testing (HALT) has been widely applied in the design phase, whose purposes are exposing the design insufficiency and potential blind failures in the early phases [103]. It should be noted that, as shown in Figure 6.1, epistemic uncertainty not only causes blind failures, but also affects the modeling and analysis of conscious failures. For example, due to lack of knowledge, the model assumption, structure, and parameters could all be affected by uncertainty [160].

In the current reliability researches, the approaches for modeling and analyzing conscious failure has been extensively discussed. However, the effect of epistemic uncertainty, especially its impact on blind failures, was still not well considered when reliability is quantified. In the following sections, based on the conceptual framework in Figure 6.1, we are going to develop approaches that support quantifying reliability/risk based on a holistic picture of its contributing factors.

6.3 Belief reliability

In this section, we introduce a new metric of reliability, belief reliability, to explicitly account for the influence of epistemic uncertainty on model-based reliability methods. The belief reliability is developed based on model-based reliability methods (e.g., structural reliability models), which is reviewed in Sect. 6.3.1. Then, the definition of belief reliability is presented in Sect. 6.3.2. In Sect. 6.3.3, we present how to evaluate the epistemic uncertainty by assessing the strength of knowledge supporting the reliability assessment. This section was previously published in [158]. For more details, readers could consult the paper directly.

6.3.1 Performance margins

For a general description of model-based reliability methods, we introduce the concepts of performance parameter and performance margin:

Definition 1 (Performance parameter)**.** *Suppose failure occurs when a parameter p reaches a threshold value p_{th}. Then, the parameter p is referred to as a performance parameter, while the threshold value p_{th} is referred to as the functional failure threshold associated with p.*

According to *Definition* 1, performance parameters and functional failure thresholds define the functional requirements on a system or a component, for which three categories exist in practice:

1. Smaller-the-better (STB) parameters: if failure occurs when $p \geq p_{th}$, then, the performance parameter p is a STB parameter.

2. Larger-the-better (LTB) parameters: if failure occurs when $p \leq p_{th}$, then, the performance parameter p is a LTB parameter.

3. Nominal-the-better (NTB) parameters: if failure occurs when $p \leq p_{th,L}$ or $p \geq p_{th,U}$, then, the performance parameter p is a NTB parameter.

Definition 2 (Performance margin)**.** *Suppose p is a performance parameter and p_{th} is its associated functional failure threshold; then,*

$$m = \begin{cases} \dfrac{p_{th} - p}{p_{th}}, & \textit{if } p \textit{ is STB}, \\ \dfrac{p - p_{th}}{p_{th}}, & \textit{if } p \textit{ is LTB}, \\ \min\left(\dfrac{p_{th,U} - p}{p_{th,U}}, \dfrac{p - p_{th,L}}{p_{th,L}}\right), & \textit{if } p \textit{ is NTB} \end{cases} \tag{6.1}$$

is defined as the (relative) performance margin associated with the performance parameter p.

Remark 1. *From Definition 2, performance margin is a unitless quantity and failure occurs whenever $m \leq 0$.*

In the model-based reliability methods, it is assumed that the performance margin can be described by a deterministic model, which is derived based on knowledge of the functional principles and failure mechanisms of the component [154, 30]. Conceptually, we assume that the performance margin model has the form

$$m = g_m(\mathbf{x}), \tag{6.2}$$

where $g_m(\cdot)$ denotes the deterministic model which predicts the performance margin and **x** is a vector of input variables.

In the design and manufacturing processes of a product, there are many uncertain factors influencing the input **x** of Eq. (6.2). Thus, the values of **x** may vary from product to product of the same type. Usually, this product-to-product variability is described by assuming that **x** is a vector of random variables with given probability density functions. Then, m is also a random variable and reliability R_p is defined as the probability that m is greater than

zero. The subscript p is used to indicate that R_p is a probability measure. Given the probability density function of $\mathbf{x}$, denoted by $f_X(\cdot)$, R_p can be calculated by:

$$R_p = Pr\left(g_m(\mathbf{x}) > 0\right) = \int \cdots \int_{g_m(\mathbf{x})>0} f_X(\mathbf{x}) \mathrm{d}\mathbf{x}. \tag{6.3}$$

6.3.2 Definition of belief reliability

Belief reliability is defined in this subsection to explicitly account for the effect of epistemic uncertainty in model-based reliability methods. For this, we first define design margin and Aleatory Uncertainty Factor (AUF):

Definition 3 (Design margin)**.** *Suppose the performance margin of a component or a system can be calculated by* (6.2)*. Then, design margin m_d is defined as*

$$m_d = g_m(\boldsymbol{x}_N), \tag{6.4}$$

where $\boldsymbol{x}_N$ is the nominal values of the parameters.

Definition 4 (Aleatory Uncertainty Factor (AUF))**.** *Suppose R_p is the probabilistic reliability calculated from the performance margin model using* (6.3)*. Then, AUF σ_m is defined as*

$$\sigma_m = \frac{m_d}{Z_{R_p}}, \tag{6.5}$$

where Z_{R_p} is the value of the inverse cumulative distribution function of a standard normal distribution evaluated at R_p.

Further, let equivalent design margin M_E to be

$$M_E = m_d + \epsilon_m, \tag{6.6}$$

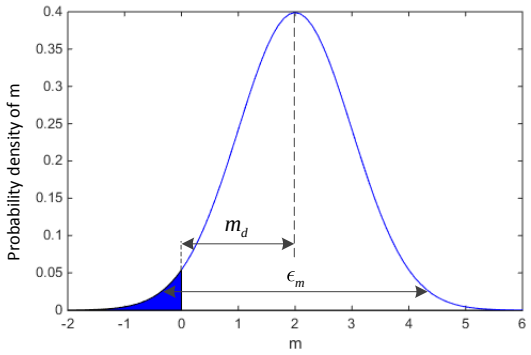
where $\epsilon_m \sim$ Normal$(0, \sigma_m^2)$. It is easy to verify that $M_E \sim$ Normal(m_d, σ_m^2) and R_p can be calculated as the probability that $M_E > 0$, as shown in Figure 6.2 (a). Therefore, the probabilistic reliability can be quantified by the equivalent performance margin and further by m_d and σ_m, where

- m_d describes the inherent reliability of the product when all the input variables take their nominal values. Graphically, it measures the distance from the center of the equivalent performance margin distribution to the boundaries of the failure region, as shown in Figure 6.2 (a);
- σ_m accounts for the uncertainty resulting from the product-to-product random variations, *e.g.* the tolerance of manufacturing processes, the variability in material properties, *etc*. Usually, these random variations are controlled by engineering activities such as tolerance design, environmental stress screening, stochastic process control, *etc* [155].

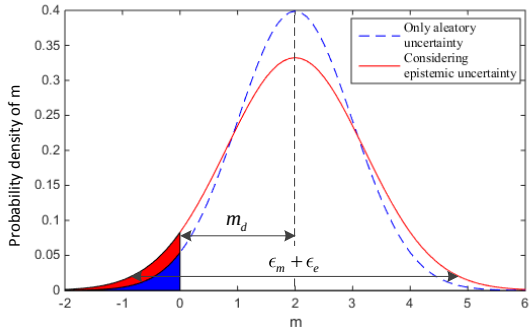
To further account for the effect of epistemic uncertainty, it is assumed that:

$$M_E = m_d + \epsilon_m + \epsilon_e, \tag{6.7}$$

where ϵ_e is an adjustment factor [169] and $\epsilon_e \sim$ Normal$(0, \sigma_e^2)$. Parameter σ_e is defined as Epistemic Uncertainty Factor (EUF) and it quantifies the effect of epistemic uncertainty. The physical meaning of (6.7) is explained in Figure 6.2 (b): epistemic uncertainty introduces additional dispersion to the aleatory distribution of the equivalent performance margin. The degree of the dispersion is related to the knowledge we have on the failure process of the product, *i.e.*, the more knowledge we have, the less value σ_e takes.



(a) Aleatory distribution



(b) Effect of epistemic uncertainty

Figure 6.2: Epistemic uncertainty effect on the distribution of the equivalent performance margin

Considering the assumption made in (6.7), we can, then, define the belief reliability as follows:

Definition 5 (Belief reliability)**.** *The reliability metric*

$$R_B = \Phi_N \left(\frac{m_d}{\sqrt{\sigma_m^2 + \sigma_e^2}} \right) \tag{6.8}$$

is defined as belief reliability, where $\Phi_N(\cdot)$ is the cumulative distribution function of a standard normal random variable.

Belief reliability can be interpreted as our belief degree on the product reliability, based on the knowledge of design margin, aleatory uncertainty and epistemic uncertainty. In the following, we discuss respectively how design margin, aleatory uncertainty and epistemic uncertainty influence the value of belief reliability.

Discussion 1. *It is obvious from* (6.8) *that $R_B \in [0, 1]$, where*

- $R_B = 0$ *indicates that we believe for sure that a component or system is unreliable,* i.e.*, it cannot perform its desired function under stated time period and operated conditions.*
- $R_B = 1$ *indicates that we believe for sure that a component or system is reliable,* i.e.*, it can perform its desired function under stated time period and operated conditions.*
- $R_B = 0.5$ *indicates that we are most uncertain about the reliability of the component or system [91].*
- $R_{B,A} > R_{B,B}$ *indicates that we believe that product A is more reliable than product B.*

Discussion 2 (Variation of R_B with the design margin)**.** *From* (6.8)*, it is easy to see that R_B is an increasing function of m_d, as illustrated by Figure 6.3, which is in accordance with the intuitive fact that when the design margin is increased, the component or system becomes more reliable.*

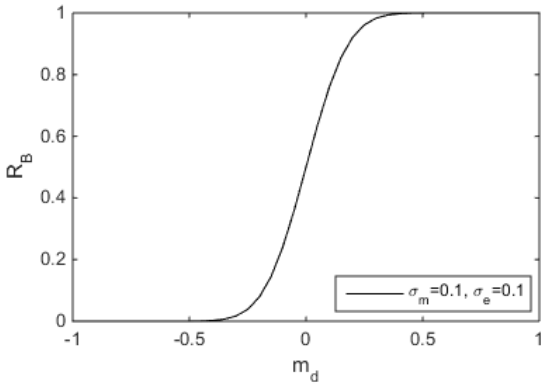


Figure 6.3: Influence of m_d on R_B

Besides, it can be verified from (6.8) *that if $m_d = 0$, $R_B = 0.5$. This is because when $m_d = 0$, the product is at borderline between working and failure. Therefore, we are most uncertain about its reliability (For details, please refer to the maximum uncertainty principle in [91]).*

Discussion 3 (Variation of R_B with the aleatory uncertainty)**.** *In* (6.8)*, the effect of aleatory uncertainty is measured by the AUF, σ_m. Figure 6.4 shows the variation of R_B with σ_m, when σ_e is fixed, for different values of m_d. It can be*

seen from Figure 6.4 that when m_d and σ_e are fixed, R_B approaches 0.5 as σ_m increases to infinity. The result is easy to understand, since $\sigma_m \to \infty$ indicates the fact that uncertainty has the greatest influence.

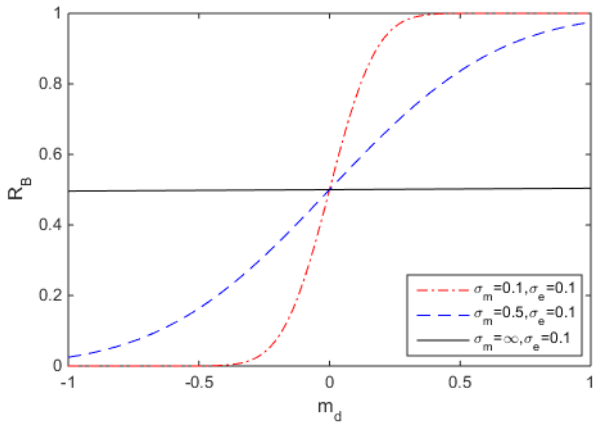


Figure 6.4: Variation of R_B with σ_m

Discussion 4 (Variation of R_B with the epistemic uncertainty)**.** *In* (6.8)*, the effect of epistemic uncertainty is measured by the EUF, σ_e. The variation of R_B with respect to σ_e is illustrated in Figure 6.5, with σ_m fixed to 0.2. From Figure 6.5, we can see that when $\sigma_e \to \infty$, R_B also approaches 0.5, for the same reason as the AUF.*

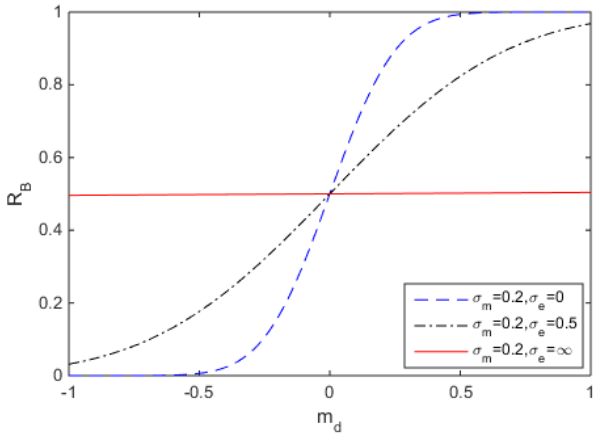


Figure 6.5: Variation of R_B with σ_e

Besides, it can be shown from (6.8) *and assumption (3) that as $\sigma_e \to 0$, R_B approaches the R_p calculated by the model-based reliability methods using equation* (6.3)*. This is a natural result since $\sigma_e = 0$ is the ideal case for which there is no epistemic uncertainty, so that the product failure behavior is accurately predicted by the deterministic performance margin model and the aleatory uncertainty.*

In practice, we always have $m_d \geq 0$ and $\sigma_e > 0$. Therefore,

$$R_B \leq R_p \tag{6.9}$$

where R_p is the probabilistic reliability predicted by (6.3) *under the same conditions. Equation* (6.9) *shows that using belief reliability yields a more conservative evaluation result than using the probabilistic reliability, because belief reliability considers the effect of insufficient knowledge on the reliability evaluations.*

6.3.3 Quantification of epistemic uncertainty

In this section, we present a method to quantify epistemic uncertainty in reliability assessment by assessing the strength of knowledge supporting the assessment. First, we discuss how to evaluate the state of knowledge.

Evaluation of the state of knowledge

In the life cycle of a component or system, the knowledge on the products' failure behavior is gained by implementing a number of engineering activities of reliability analysis, whose purposes are to help designers better understand potential failure modes and mechanisms. In this section, we refer to these engineering activities as epistemic uncertainty-related (EU-related) engineering activities. Table 6.1 lists some commonly encountered EU-related engineering activities and discusses their contributions to gaining knowledge and reducing epistemic uncertainty, where FMECA stands for Failure Mode, Effect and Criticality Analysis, FRACAS stands for Failure Reporting, Analysis, and Corrective Action System, RET stands for Reliability Enhancement Test, RGT stands for Reliability Growth Test and RST stands for Reliability Simulation Test.

Table 6.1: Examples of EU-related engineering activities

Activities	Contributions to gaining knowledge and reducing epistemic uncertainty
FMECA	FMECA helps designers to identify potential failure modes and understand their effects, so as to increase the designer's knowledge about potential failures [25].
FRACAS	By implementing FRACAS, knowledge on potential failure modes and mechanisms is accumulated based on previously ocurred failures and corrective actions [22].
RGT	In a RGT, cycles of Test Analysis and Fix (TAAF) are repeated until the product reaches its reliability requirements. In this way, designers' knowledge on the failure modes and mechanisms is accumulated [142].
RET	As the RGT, RET reduces epistemic uncertainty by stimulating potential failures, but using highly accelerated stresses, which can generate failures that are hard to be identified by analyses or conventional tests [22].
RST	In a RST, simulation tests are conducted based on physics-of-failure models to identify weak design points for the products. Knowledge of potential failure modes can be accumulated in this way [109**?**].

In this paper, we make an assumption that the state of knowledge is directly related to the effectiveness of the EU-related engineering activities. Suppose there are n EU-related engineering activities in a product life cycle. Let $y_i, i = 1, 2, \cdots, n$ denote the effectiveness of the EU-related engineering activities, where $y_i \in [0, 1]$; the more effective the engineering activity is, the larger value the corresponding y_i takes. The values of y_i are determined by asking experts to evaluate the effectiveness of the EU-related engineering activities, based on a set of predefined evaluation criteria. Examples of the evaluation criteria and how to use them to determine the values of y_i can be found in [158].

Determination of EUF

Having determined the value of y, we need to define a function $\sigma_e = h(y)$, through which σ_e is determined. Since σ_e is a measure of the severity of epistemic uncertainty and y measures the state of knowledge, σ_e is negatively dependent on y. Theoretically, any monotonic decreasing function of y could serve as $h(y)$. In practice, the form of $h(y)$ reflects the decision maker attitude towards epistemic uncertainty and is related to the complexity of the product. Therefore, we propose $h(y)$ to be

$$h(y) = \begin{cases} \dfrac{1}{3\sqrt{y}} \cdot m_d, & \text{for simple products;} \\ \dfrac{1}{3y^6} \cdot m_d, & \text{for complex products;} \\ \dfrac{1}{3y^2} \cdot m_d, & \text{for medium complex products.} \end{cases} \tag{6.10}$$

By letting $\sigma_m = 0$ and m_d fixed to a constant value, the attitudes of the decision maker for different products can be investigated (see Figure 6.6):

- for simple products, R_B is a convex function of y, indicating that even when y is small, we can gather enough knowledge on the product function and failure behaviors, so that we can assign a high value to the belief reliability;
- for complex products, R_B is a concave function of y, indicating that only when y is large we can gather sufficient knowledge on the product function and failure behaviors, so that we can assign a high value to the belief reliability;
- the $h(y)$ for medium complex products lies between the two extremes.

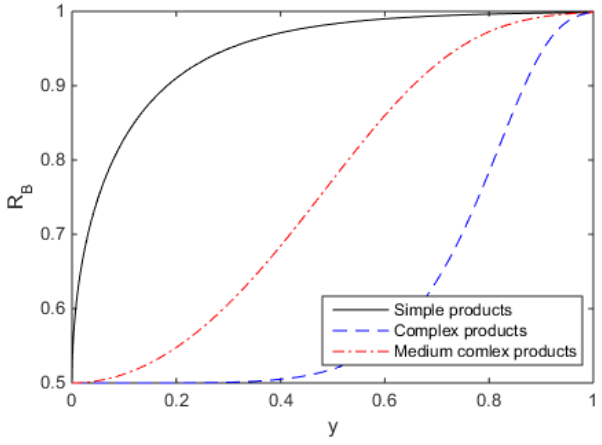


Figure 6.6: Different attitudes of the decision maker towards epistemic uncertainty

6.4 Belief risk index

Based on the conceptual model of failures in Sect. 6.1, we also extend the traditional risk index to integrate the influence of epistemic uncertainty. Since in the traditional probabilistic risk assessment models, the performance margins are not calculated explicitly, we develop an uncertainty equivalence model first to reconstruct performance margins based on the risk indexes calculated by the PRA (Sect. 6.4.1). Then, the belief risk index is defined in Sect. 6.4.2. The work in this section was previously published in [160]. More details can be found in the original paper.

6.4.1 Uncertainty equivalence model

Let us consider a generic PRA model:

$$Risk_P^* = f(\boldsymbol{\theta}^*) \tag{6.11}$$

In Eq. (6.11), $Risk_P^*$ is the estimated probabilistic risk index ($*$ indicates estimates); $f(\cdot)$ denotes the PRA model used for calculating $Risk_P^*$ (e.g., fault trees, event trees); $\boldsymbol{\theta}^*$ is a vector containing the estimated values of the model parameters. $Risk_P^*$ is an estimate of the true (but unknown) frequency of the given consequence, based on the model $f(\cdot)$ and the estimates $\boldsymbol{\theta}^*$ of the true (but unknown) values of the model parameters.

An uncertainty equivalence model is developed to integrate the epistemic uncertainty in the PRA with the (aleatory) uncertainty in the model prediction ($Risk_P^*$). For this, let us first artificially construct a PRA model for the same consequence of Eq. (6.11), where the safety margin of the corresponding consequence is S_E, i.e., the consequence occurs whenever $S_E < 0$. Please note that here we use safety margin as the risk assessment community is more familiar with this term than performance margin. However, safety margin and performance margin are considered equivalent by us, and we use these two terms interchangeably. Let us further assume that:

Assumption 1. *The safety margin S_E follows a normal distribution $S_E \sim \left(\mu_{S_E}, \sigma_{S_E}^2\right)$ and*

$$\frac{\mu_{S_E}}{\sigma_{S_E}} = -\Phi^{-1}\left(Risk_P^*\right) \tag{6.12}$$

where $Risk_P^*$ is calculated by the PRA model in Eq. 6.11 and $\Phi^{-1}(\cdot)$ is the inverse cumulative distribution function of a standard normal distribution.

From Eq. (6.12), it is easy to verify that

$$\Pr\left(S_E < 0\right) = \Phi\left(-\frac{\mu_{S_E}}{\sigma_{S_E}}\right) = Risk_P^*, \tag{6.13}$$

where $\Phi(\cdot)$ is the cumulative distribution function (CDF) of a standard normal distribution. Hence, the uncertainty in the predicted risk index $Risk_P^*$ is equivalent to the artificially constructed PRA model with a safety margin S_E .

Therefore, the auxiliary random variable S_E is called equivalent safety margin.

Please note that the purpose of making Assumption 1 is to artificially construct a PRA with equivalent uncertainty as the original PRA model in Eq. (6.11). For this, the equivalent safety margin S_E only needs to satisfy $\Pr(S_E < 0) = Risk^*$: that is, any random variable that satisfies $\Pr(S_E < 0) = Risk^*$ can be selected as the equivalent safety margin. However, to integrate epistemic uncertainty in the developed model, one also needs to identify the distribution of the equivalent safety margin. The selected distribution should reflect the decision makers' prior belief on how the safety margin is distributed and will be updated by integrating epistemic uncertainty. In practice, the distribution of the equivalent safety margin can be determined through expert elicitation . In Assumption 1, for simplicity and illustrative purposes, we directly assume that the equivalent safety margin follows a normal distribution. The developed methods, however, can be naturally extended to other distributions.

It should also be noted that even though we adopt Assumption 1, there are still infinite choices of μ_{S_E}, σ_{S_E}, as long as Eq. holds. In practice, we can fix one of the two parameters and calculate the other one from Eq. . It can be seen in Sect. 4.2 that the values of μ_{S_E}, σ_{S_E} do not affect the value of the belief risk index, provided that their values satisfy Eq. . For example, in the illustrative example of event tree models, as $Risk_P^* = 10^{-3}$, if we set $\sigma_{S_E} = 1$, then, from Eq. , $\mu_{S_E} = -1 \cdot \Phi^{-1}(Risk_P^*) \cdot \sigma_{S_E} = 3.0902$. Hence, the uncertainty in the result of the event tree analysis can be viewed as equivalent to an artificially constructed PRA, where the equivalent safety margin is $S_E \sim (3.0902, 1)$.

To integrate EU in the uncertainty equivalence model, we make the same assumption as that in Figure 6.2, i.e., EU increases the dispersion of the distribution of the equivalent safety margin but does not affect its center.

Therefore, the effect of epistemic uncertainty can be modeled by replacing the equivalent safety margin S_E with the EU-affected equivalent safety margin $S_E^{'}$:

$$S_E^{'} = S_E + \epsilon_e, \tag{6.14}$$

where ϵ_e is the adjustment factor for EU and is assumed to be

$$\epsilon_e \sim \left(0, \sigma_e^2\right). \tag{6.15}$$

Eq. (6.14) shows that by making Assumptions 1 and 2, the overall uncertainty (including the uncertainty in $Risk_P^*$ and the EU) in the PRA is equivalent to that of presumed PRA with a safety margin $S_E^{'}$. Hence, the model in Eq. (6.14) is called uncertainty equivalence model. The parameter σ_e controls how much EU affects the results of the PRA and is assumed to be proportional to the mean μ_{S_E} :

$$\sigma_e = \alpha_e \cdot \mu_{S_E}. \tag{6.16}$$

The parameter α_e characterizes the magnitude of the effect of EU on the PRA model results and is called

Epistemic Uncertainty Factor (EUF). The EUF takes values in $[0, \infty]$ where a large value indicates a large effect of EU. The meaning of the EUF parameter is the additional dispersion brought by EU on the equivalent safety margin. Its evaluation takes into account the impact of epistemic uncertainty and will be discussed in subsequent sections.

6.4.2 Definition of belief risk index

A new risk index, called belief risk index ($Risk_B$), is, then, defined based on the uncertainty equivalency model, to consider the effect of EU on PRA:

$$Risk_B \triangleq \Pr\left(S_E^{'} < 0\right) = \Phi\left(-\frac{\mu_{S_E}}{\sqrt{\sigma_{S_E}^2 + \sigma_e^2}}\right). \tag{6.17}$$

It can be seen from Eq. (6.17) that the belief risk index is the probability that the (EU) affected safety margin is less than zero. The concept of safety margin has been widely used in structural reliability analysis . It can be seen that when there is no epistemic uncertainty and the safety margin follows a normal distribution, the belief risk index as defined in Eq. (6.17) is equivalent to the structural unreliability. The major difference between our developed metric in Eq. (6.17) and the traditional structural reliability theory is that, Eq. (6.17) allows explicitly considering epistemic uncertainty, which is not considered in the structural reliability theory. The belief risk index is defined with respect to a specific consequence and measures the uncertainty on the occurrence of this consequence. It should be noted that the probability here takes the subjective interpretation: it measures the belief degree on the occurrence of a given consequence, based on both the prediction of the PRA model and the EU that affects the PRA model. From Eq. (6.17), we can see that the uncertainty in $Risk_B$ is equivalent to that of a PRA model result where the safety margin is $S_E^{'} \sim \left(\mu_{S_E}, \sigma_{S_E}^2 + \sigma_e^2\right)$. This, however, does not mean that the belief risk index can be interpreted based on the frequentist interpretation of probability. Rather, the belief risk index is a subjective metric that allows comparing the decision makers' personal belief degree on the uncertainty in the PRA results. It should be noted that sometimes the order relationships indicated by the belief risk indexes is more interesting to the decision makers than the absolute values. For example, suppose that we have two cases where $Risk_{B,1} = 0.0466$ and $Risk_{B,2} = 0.0105$: a proper interpretation is that we are more confident (less uncertain) in the second case that the predicted consequence by the PRA model will not occur.

Equation (6.17) is not very easy to use in practice, as one usually only knows the value of $Risk_P^*$ (calculated from the PRA model), not μ_{S_E} and σ_{S_E}. Substituting Eqs. (6.11) and (6.16) into Eq. (6.17), we have

$$Risk_B = \Phi\left(\frac{\Phi^{-1}\left(Risk_P^*\right)}{\sqrt{1 + \left(\alpha_e \cdot \Phi^{-1}\left(Risk_P^*\right)\right)^2}}\right). \tag{6.18}$$

Eq. (6.18) facilitates the practical evaluation of the belief reliability index, as it allows calculating $Risk_B$ based on

the estimation of $Risk_P^*$.

The effect of EU on the belief risk index can be investigated graphically. First note that both S_E and $S_E^{'}$ have been assumed to follow normal distributions. Therefore, they can be transformed into the standard normal space by taking the transformation:

$$Z = \frac{X - \mu_X}{\sigma_X}, \tag{6.19}$$

where $X = S_E, S_E^{'}$, respectively.

It is easy to verify that in the standard normal space, the distance from the origin to the failure region associated with S_E is

$$d_P = \left|\Phi^{-1}\left(Risk_P^*\right)\right| = \left|\frac{\mu_{S_E}}{\sigma_{S_E}}\right|, \tag{6.20}$$

while after considering the EU, the distance becomes

$$d_B = \left|\Phi^{-1}\left(Risk_B\right)\right| = \left|\frac{\Phi^{-1}\left(Risk_P^*\right)}{\sqrt{1+\alpha_e^2 \cdot \left(\Phi^{-1}\left(Risk_P^*\right)\right)^2}}\right| = \frac{d_P}{\sqrt{1+\alpha_e^2 \cdot {d_P}^2}}, \tag{6.21}$$

as shown in Figure 6.7. As $\alpha_e \geq 0$, we have $d_B \leq d_P$, which shows that considering EU decreases the safety margin. Therefore, $Risk_B$ always provides a more conservative value of the risk index than the probabilistic risk index.

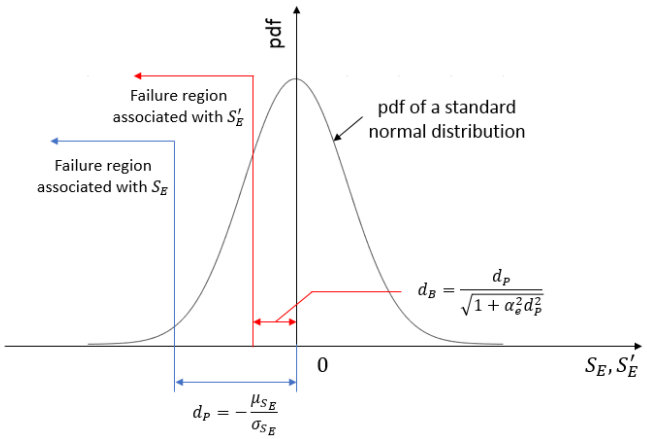


Figure 6.7: Graphical interpretation of epistemic uncertainty

Another observation is that when the EU has the most severe influence, we have $\alpha_e \to \infty$ and $Risk_B \to 0.5$. This indicates that $Risk_B = 0.5$ is a state of maximal EU: at this state, one is totally ignorant about the system state due to the influence of the EU (we cannot judge whether the consequence is more likely to occur or not to occur). Therefore, $Risk_B$ can be regarded as a measure of confidence on the result of the PRA: the closer $Risk_B$ to 0.5, the less sure one is about the result of the PRA, and, then, one should not to use the PRA model results for decision making.

6.4.3 Indifference method for belief risk index evaluation

A critical step in belief risk index evaluation is to determine the value of α_e. In [160], we proposed an approach to determine the value of α_e based on the maturity of epistemic uncertainty management for the PRA (denoted by M_{EUM}). Based on the severity of the influence of EU on the PRA results, we define five levels of M_{EUM}): Initial $(M_{EUM} = 1)$, Uncontrolled $(M_{EUM} = 2)$, Complete $(M_{EUM} = 3)$, Adequate $(M_{EUM} = 4)$ and Accurate $(M_{EUM} = 5)$, with increasing degree of maturity for managing the epistemic uncertainty. Details definitions of each level and their assessment guidelines can be found in Chapter 8 of this thesis or directly from [160].

As the value of M_{EUM} relates to the level of EU, where $M_{EUM} = 1$ means that the impact of EU is the greatest whereas $M_{EUM} = 5$ indicates the lowest impact, obviously, α_e is a decreasing function of M_{EUM} :

$$\alpha_e = h\left(M_{EUM}\right). \tag{6.22}$$

The function $h\left(\cdot\right)$ reflects the tolerance on the EU. This can be shown by investigating the dependence of $Risk_B$ on M_{EUM}, when the PRA model predicts that $Risk_P^* = 0$, as shown in Figure 6.8. Once $h\left(\cdot\right)$ is known, Figure 6.8 can be drawn by letting $Risk_P^* \to 0$:

$$\begin{aligned} Risk_B|Risk_P^* \to 0 &= \Phi\left(\frac{\Phi^{-1}\left(Risk_P^*\right)}{\sqrt{1+\alpha_e^2\cdot\left(\Phi^{-1}\left(Risk_P^*\right)\right)^2}}\right) = \Phi\left(-\frac{1}{\sqrt{\left(\frac{1}{\Phi^{-1}\left(Risk_P^*\right)}\right)^2+\alpha_e^2}}\right) \\ &= \Phi\left(-\frac{1}{\alpha_e}\right) = \Phi\left(-\frac{1}{h\left(M_{EUM}\right)}\right). \end{aligned} \tag{6.23}$$

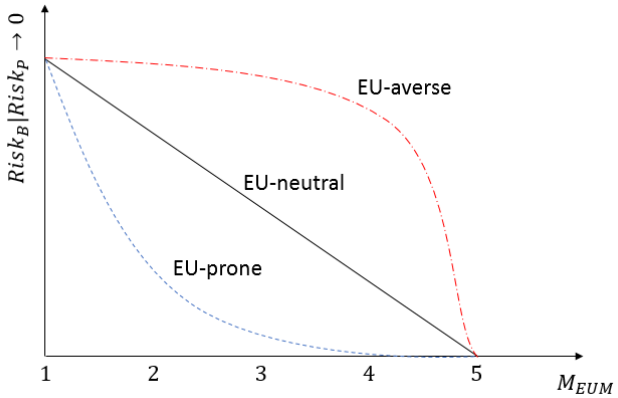


Figure 6.8: Typical behaviors of $Risk_B|Risk_P^* \to 0$ under different values of M_{EUM}.

Typically, the attitude towards EU exhibits three types of behaviors, *i.e.*, EU-averse, EU-neutral and EU-prone:

- for EU-prone, $Risk_B|Risk_P^* \to 0$ is a convex function of M_{EUM} , meaning that even though the EU is quite large (M_{EUM} is relatively immature), there is willingness to trust the prediction of the PRA model;
- for EU-averse, $Risk_B|Risk_P^* \to 0$ is a concave function of M_{EUM} , meaning that only when the EU is very small (M_{EUM} is highly mature), there is willingness to trust the prediction of the PRA model;
- EU-neutral lies between the two extremes: $Risk_B|Risk_P^* \to 0$ is approximately a linear function of M_{EUM}.

In this paper, we suggest the following form of $h(\cdot)$, for its flexibility to model EU-averse, EU-prone, and EU-neutral attitudes:

$$\alpha_e = h(M_{EUM}) = K\left(\frac{1}{M_{EUM}-1} - \frac{1}{4}\right), \tag{6.24}$$

where K is a parameter that determines the attitude towards EU, as shown in Figure 6.9.

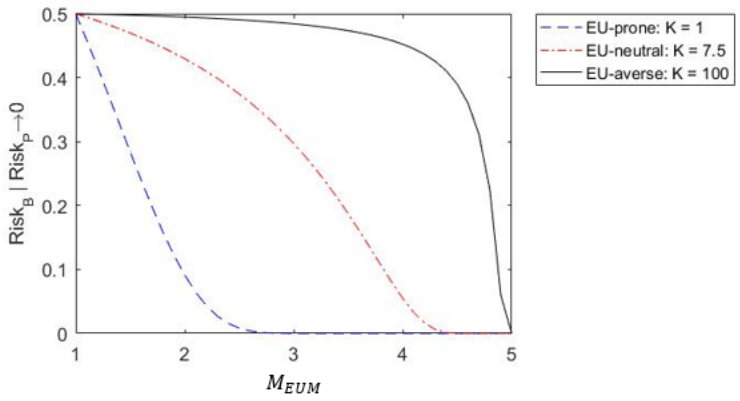


Figure 6.9: Attitude towards EU at different values of K.

The values of can be estimated using an indifference method. A survey is first conducted to collect empirical data from the decision makers. The decision makers are asked to the following thought experiment : Suppose you are concerned with a given consequence of an accident, which, if it occurs, brings you a financial loss of one Yuan (Chinese currency). An insurance company offers you an insurance plan : if the accident actually occurs, you will get a reimbursement of one Yuan. Suppose that you conduct a PRA, which shows that $Risk_P^* = 0$, *i.e.,* there is no risk on this specific consequence at all. Suppose we have five cases, where the for the PRA process is Unmanaged ($M_{EUM} = 1$), Uncontrolled ($M_{EUM} = 2$), Complete ($M_{EUM} = 3$), Adequate ($M_{EUM} = 4$) and Accurate ($M_{EUM} = 5$), respectively. Then, for the five cases, what are the highest prices that make you willing to buy the insurance, respectively?

Denote the answers from the decision makers by $\beta_1, \beta_2, \cdots, \beta_5$, respectively. These values, then, reflect the decision makers' beliefs on the values of $Risk_B|Risk_P^* \to 0$ under different levels of M_{EUM}. Therefore, they can be used as empirical data for estimating the parameter K. Note that we have $\beta_1 = 0.5$, as $M_{EUM} = 1$ is the state with maximal EU and, therefore, corresponds to $R_B = 0.5$, as shown in Sect. 6.4.1. Similarly, we have $\beta_5 = 0$, as

$M_{EUM} = 5$ is the state with no EU, when we trust the prediction of the PRA model. Therefore, the decision makers only need to assign values to β_2, β_3 and β_4.

In this paper, we use the least square method for parameter estimation, in which the value of K is found by solving the following minimization problem:

$$\min_K SSE = \sum_{i=1}^{5} \left(\beta_i - \Phi \left(-\frac{1}{K\left(\frac{1}{M_{EUM}-1} - \frac{1}{4}\right)} \right) \right), \tag{6.25}$$

where SSE represents the sum of square error of the data and $\beta_i, i = 1, 2, \cdots, 5$ are empirical data collected from experts based on the thought experiment presented before. The optimization problem can be solved easily using standard nonlinear programming methods, *e.g.*, sequential quadratic programming.

6.5 Summary of major contributions

This chapter summarizes the major findings in my research axis 1, which aims at integrating epistemic uncertainty into risk and reliability quantification. On this aspect, we achieved the following representative results:

- we developed a generic conceptual framework for understanding failure causes. The framework distinguishes between conscious failure and blind failure, and summarizes the main contributors to failures as design margin, aleatory uncertainty, degradation and epistemic uncertainty. By doing so, the contribution from epistemic uncertainty on failures can be explicitly considered.
- new indexes are developed to quantify risk and reliability, which explicitly considers the impact of knowledge and epistemic uncertainty. Compared to the existing risk/reliability metrics, the proposed metrics allows explicitly considering the impact of epistemic uncertainty. Hence, the developed metrics can better reflect the reality, especially when blind failures are the major failure sources.

Chapter 7

MODELING DEPENDENT DEGRADATION PROCESSES WITH STOCHASTIC HYBRID AUTOMATON

This chapter summarizes some of my representative research results in research axis 2. The focus of this axis is to develop a generic model and efficient analysis algorithms for dependent failure behaviors involving both discrete and continuous degradation variables. In Sect. 7.1, we briefly review the related literature and define the research problems of this research axis. Some representative results are briefly introduced in Sects. 7.2 - 7.4: In Sect. 7.2, we develop a generic modeling framework based on Stochastic Hybrid Automaton (SHA) for dependent failure behaviors; in Sect. 7.3, a semi-analytical method reliability assessment framework is developed to improve the computational efficiency; in Sect. 7.4, Common Cause Failure (CCF) is modeled and analyzed based on the developed frameworks in Sect. 7.4. Finally, in Sect. 7.5, we summarize the major contributions achieved in this research axis.

7.1 Research questions

Failure of industrial components, systems and products may be caused by multiple failure processes, e.g. wear, corrosion, erosion, creep, fatigue, etc. [70]. In general, the failure processes are categorized as degradation processes (or soft failures) and catastrophic failure processes (or hard failures) [85]. Soft failure is caused by continuous degradation and is often modeled by a continuous-state random process, e.g., Wiener process, Gamma process, inverse Gaussian process, continuous-time semi Markov process, etc. Hard failure is caused by traumatic shocks in various patterns and is often modeled by a discrete-state random process, e.g., Homogeneous Poisson

Process (HPP), Nonhomogeneous Poisson Process (NHPP), etc. Often, complex dependencies exist among the failure processes [156]. For example, [74] presents experimental data to show that erosion and corrosion can enhance each other and therefore accelerate the failure process. Also, it is observed in [97] that the dependency between creep and fatigue severely reduces the Time-To-Failure (TTF) of the specimens that are exposed to high temperatures and heavy loads. How to accurately model the failure behavior resulting from the interdependent degradation (continuous) and shock (discrete) processes, is, therefore, an important question in reliability modeling.

In literature, various methods have been developed to to consider the dependent failure behavior among degradation processes and random shocks. For example, Peng et al. [111] develop a dependency model where the arrived shocks lead to an abrupt increase of the degradation process. Wang and Pham [136] investigate systems subject to dependent competing risk, which suffer failures due to degradations and random shocks: the model is proposed of shocks that can cause immediate failure of the system, with a time-dependent probability $p(t)$, or can increase the degradation level with probability $(1 - p(t))$. Cha and Finklstein [27] assume that a shock can lead to a hard failure with probability $p(t)$, or can increase the degradation rate with probability $(1 - p(t))$. Jiang et al. [69] develop a model that considers that the threshold of hard failures can be shifted by random shocks. Rafiee et al. [117] consider that the degradation rate is increased by a series of shocks. Jiang et al. [70] categorize shocks into different shock zones based on their magnitudes and consider that shocks in different zones have different effects on the degradation process. Bagdonavicius et al. [16], Fan et al. [40] and Ye et al. [145] develop models that consider that the probability of hard failures is increased as the degradation process progresses. Huynh et al. [63] investigate maintenance strategies for a dependence model, where the intensity of the NHPP for random shock is a piecewise function of the degradation magnitude.

Although a substantial amount of works have been done, as reviewed above, two issues still remain to be addressed:

1. In most of the existing works, the modeling process varies a lot depending on the context of the application and the resulted models are also highly case-specific. No generic model and modeling approaches are available for a general dependent degradation-shock process.

2. To calculate the reliability, most of the existing approaches rely on Monte Carlo simulation, which brings large amount of computational burden. Efficient algorithms, is, therefore, also needed to reduce the computational burden of the reliability assessment.

In this chapter, we focus on these two research questions. Section 7.2 propose a generic modeling framework for dependent degradation-shock processes for the first research question. Sections 7.3 and 7.4 focus on the second research question, in which we propose a semi-analytical approach for efficient reliability assessment and apply it to a general dependent degradation-shock process (Sect. 7.3) and system CCF with dependent degradation-shock failure processes in its components (Sect. 7.4).

7.2 A generic framework for modeling continuous and discrete degradation with dependencies

In this section, we develop a generic modeling framework for dependent degradation-shock processes based on stochastic hybrid automation (SHA). The framework is presented in Sect. 7.2.1 and applied on a real-world case study in Sect. 7.2.2. This section is based on our publications [41] and [42]. For more details, readers could refer to these works.

7.2.1 The framework

To address the first research problem in Sect. 7.1, we develop a generic framework, based on SHA, for modeling the dependent failure behavior involving both continuous and discrete degradation. SHA is a widely used model for describing system behaviors that are stochastic and are hybrid (mix of discrete and continuous transitions) [26]. It has been used to model complex system behaviors, such as bio-chemical systems, collision detection in aviation, analysis of telecommunication systems, etc. As illustrated in Figure 7.1, an SHA can be represented by a tuple: $SHA = (Q, E, X, A, A_C, H, F, P, q_0, x_0, P_0)$, where

- $Q = [q_1, q_2, \cdots, q_n]$ and $X = [x_1, x_2, \cdots, x_c]$ represent discrete and continuous states, respectively;
- $E = [e_1, e_2, \cdots, e_r]$ represent trigger events, whose occurrence would lead to transitions among the discrete states;
- $A = [A_{i,j}]$ is a set containing all the possible transitions between the discrete states i and j. The element in A, $A_{i,j}$ is further defined by $A_{i,j} = [q_i, q_j, e_{i,j}, G_{i,j}, R_{i,j}]$, where $e_{i,j}$ is the trigger event for this transition, $G_{i,j}$ is the gate condition that prevents the transition from happening, and $R_{i,j}$ is a reset map that resets the values of the continuous variables after the transition.
- $A_c : X \times Q$ defines how do the continuous variables change over time. Depending on the application, A_c could take different forms like difference equation, differential equations, stochastic differential equation, etc.
- H and F are counters used for tracking the transition time and its PDF;
- P is a matrix that defines the transition probabilities among the different states.

Abstractly, a dependent failure model can be viewed in terms of three parts:

- continuous processes, which are typically used to model the continuous degradation, and other phenomena like continuous variation of environmental parameters;

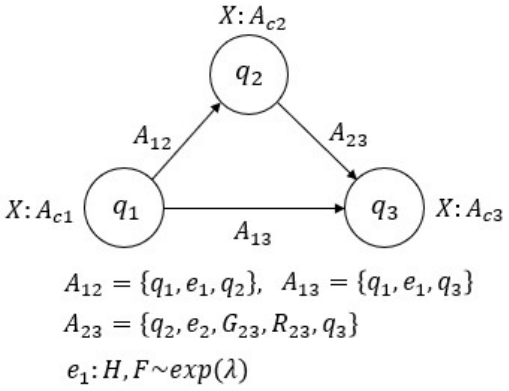


Figure 7.1: A graphical illustration of SHA.

- discrete events, which can be used to describe abrupt changes to the degradation processes, caused by factors like failure of components, change of control laws, etc., or the degradation process that are itself discrete in nature;
- dependency relations, which describes how do continuous process and discrete events are influenced by one another.

The three parts of a generic degradation model can be naturally captured by an SHA model. In Figure 7.2, we summarize how to model a generic dependent failure process using SHA. Using the framework in Figure 7.2, one is able to transform a dependent degradation process into an SHA model. Then, the behaviors of the SHA-based degradation model can be easily simulated, as many simulation software is available for simulating the behaviors of SHA, e.g. Matlab Simulink, Scilab. The reliability can, then, be determined based on the values of the degradation variables.

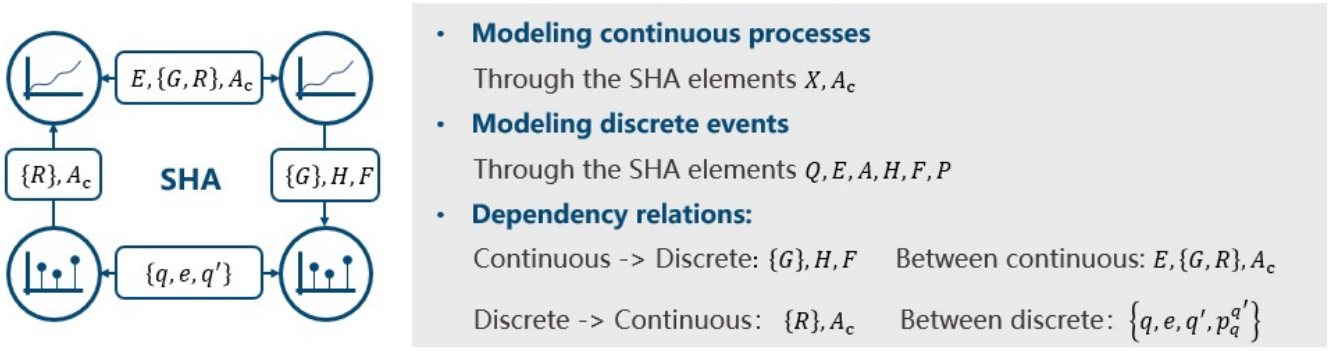


Figure 7.2: Modeling dependent degradation processes based on SHA.

We illustrate the use of the developed framework in 7.2 through benchmark example from literature [111]. A MEMS system is subject to two competing failure processes [111]:

- a continuous degradation process caused by failure mechnisms like wear. The degradation measure grows

following $dx = f(t;\theta)dt$. The degradation threshold is D, i.e., failure occurs when $x > D$.

- a shock process which caused by random environmental effect and is modeled by a homogeneous Poisson process with a rate λ:
  - if the shock is not fatal, some extra damage d is caused to the degradation process (with probability p_d);
  - if the shock is fatal, (p_f), the MEMS fails directly.

The two failure processes are illustrated graphically in Figure 7.3. The dependency between the two failure processes lies in the fact that a non-fatal shock brings an additional damage to the degradation process, as shown in Figure 7.3 (a).

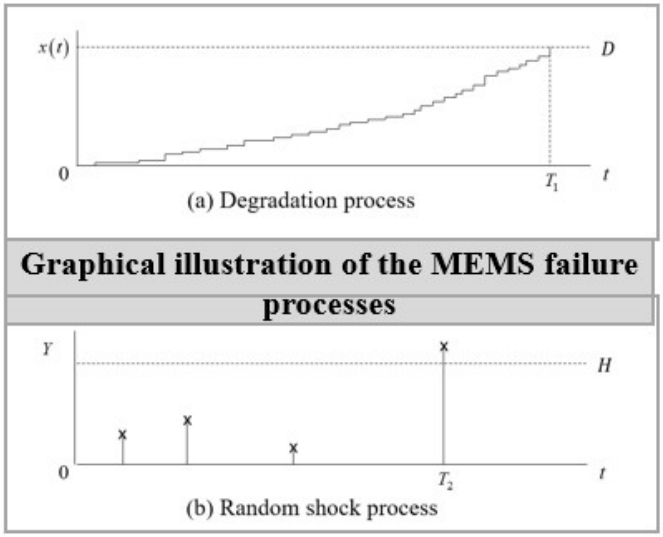


Figure 7.3: Graphical illustration of the MEMS failure processes [111]

Following the modeling framework in Figure 7.2, an SHA model can be developed for the failure processes of the MEMS. The resulted SHA model is given in Figure 7.4. In this model, state q_1 represents normal operation state, while state q_2 represents the failure state. Variable x is the degradation indicator and it degrades following the law in A_{C1}. The transition A_{11} is triggered by the arrival of a non-fatal shock, which is characterized by the clock H, F and the probability p_d. When transition A_{11} occurs, the reset map R_{11} applies, which models the additional increment brought by the shock. The transition A_{12} models the arrival of a fatal shock.

7.2.2 An application on an aviation sliding spool valve

To further test the applicability of the developed modeling framework, we apply it on a real-world case study of an aviation sliding spool. Sliding spools are critical control components in hydraulic control systems. As illustrated in Figure 7.5, a sliding spool is composed of a spool and a sleeve, where the spool slides in the sleeve to control hydraulic oil flows [144].

$$SHA = (Q, E, X, A, A_c, H, F, P, q_0, x_0, P_0)$$

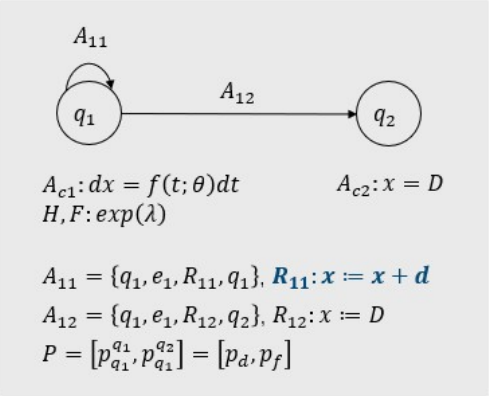


Figure 7.4: SHA model for the MEMS failure processes.

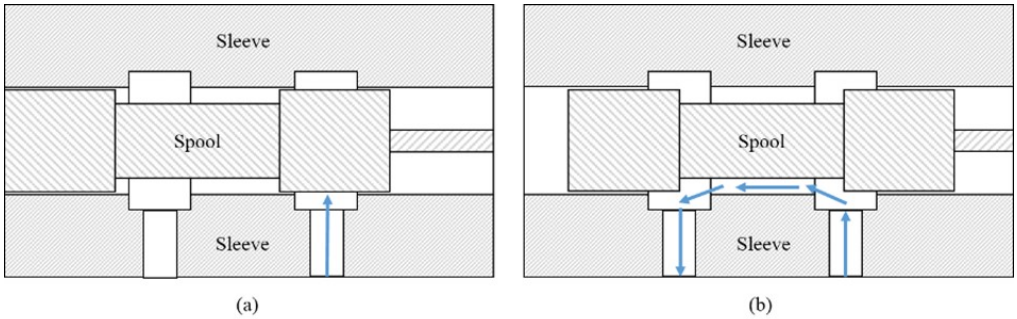
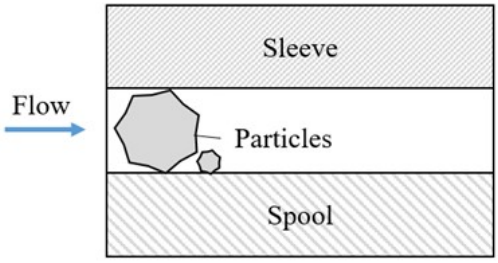


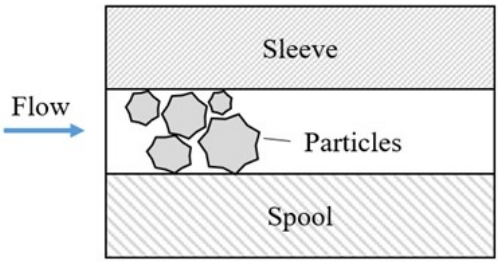
Figure 7.5: Illustration to a sliding spool: (a) Closed position; (b) Open position [144]

A sliding spool is subject to two failure mechanisms [89]. One is wear between the spool and the sleeve, and the other is clamping stagnation (also referred to as hydraulic locking or sticking), in which the spool is stuck in the sleeve. In practice, wear can be modeled by a physical deterioration model, for example, the linear Archard model suggested by Liao [89]. The modeling of clamping stagnation, however, is more complex because the factors contributing to clamping stagnation are much more varied. According to the survey by Sasak and Yamamot [124], one of the major causes of clamping stagnation is the sudden appearance of pollutant in the hydraulic oil, which can be modeled by a random shock model. The pollutant may come inside the hydraulic system from the outside environment or be generated by the hydraulic system itself. One significant source of internal pollutant is the wear debris generated due to the wear of the sliding spool. Therefore, the random shock process is dependent on the wear process: As the wear process progresses, more wear debris is generated. The debris will contaminate the hydraulic oil and further, increase the likelihood of clamping stagnation [114]. This kind of dependence, caused by the influence of degradation on shocks, needs to be considered when developing the reliability model of the spool valve. Furthermore, experimental results show that the most harmful effect of shocks is generated by wear debris, whose sizes are either close to or much smaller than the clearance of the sliding spool [124]. This is because the clamping stagnation is caused by two failure mechanisms, immediate stagnation and cumulative stagnation

[167]. When a particle with a size close to the clearance is generated, it causes immediate stagnation of the sliding spool [167], as shown in Figure 7.6 (a). If particle sizes are smaller than the clearance, the particles can enter the clearance with the hydraulic oil and form filer cakes cumulatively [166]. When the filter cakes become large enough, cumulative stagnation occurs, as shown in Figure 7.6 (b). According to Zhou [167], particles whose sizes are greater than the clearance have little effect on clamping stagnation, because they are blocked outside the clearance. Hence, based on their magnitudes, the random shocks affecting clamping stagnation can be classified into three zones with different effects on the clamping stagnation. In other words, the sliding spool is subject to zone shocks.



(a) Immediate stagnation



(b) Cumulative stagnation

Figure 7.6: Two failure mechanisms leading to clamping stagnation

More formally, the failure of the sliding spool comprises two failure processes: soft failures due to a degradation process (Figure 7.7 (a)) and hard failures due to random shocks. According to their magnitudes, random shocks are divided into three zones: damage zone, fatal zone and safety zone [70]. As shown in Figure 7.7 (b-1), a shock in the damage zone generates a cumulative damage to the system; as shown in Figure 7.7 (b-2), a shock in the fatal zone causes immediate failure of the system; a shock in the safety zone has no effect on the system's failure behavior. Note that the three shock zones in Figure 3 are shown for illustrative purposes only. In different applications, case-specific shock zones can be defined based on the magnitude of the shocks.

Degradation-shock dependence exists among the failure processes, that is, the arrival rate of the random shocks is dependent on the degradation levels, as illustrated in Figure 7.7. Failures occur whenever one of the three events happens:

- the degradation process reaches its threshold T_1;

- the cumulative damage resulting from the shocks in the damage zone exceeds its threshold T_2;
- a shock in the fatal zone occurs T_3.

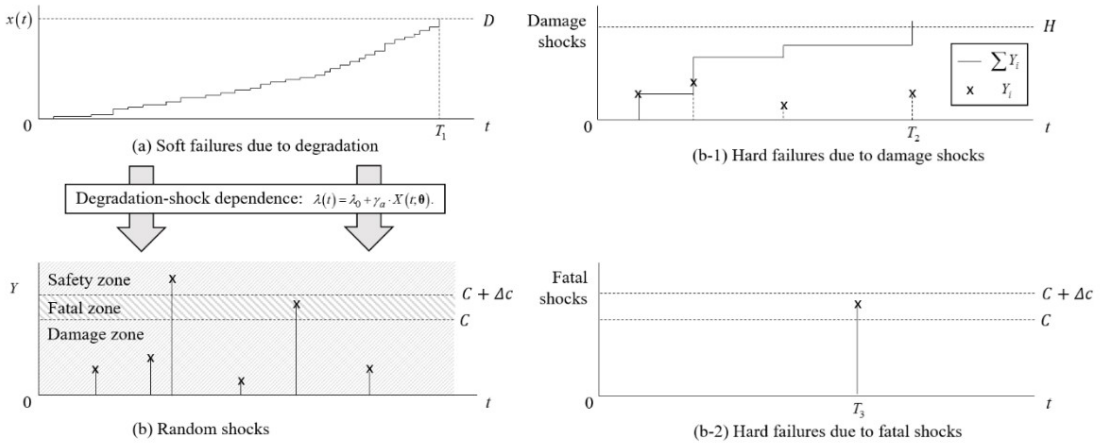


Figure 7.7: Soft failures due to degradation and hard failures due to random shocks

Based on the developed modeling framework, an SHA model can be derived for the sliding spool, as shown in Figure 7.8. Two discrete states are used to model the behavior of the sliding spool, where state 1 represents normal operation while state 2 represents failure state. The continuous variables include x, which represents the wear depth caused by the sliding wear process, and ΣW, which represents the accumulative increments of the wear depth brought be the shock processes. Transition A_{11} models the evolution of the wear depth, where the gate condition G_{11} describes the arrival of a shock in the damage zone and the reset map R_{11} models the additional increment to the wear depth caused by the shock. Transition A_{22} models the arrival of the fatal shocks. Reliability analysis can be done by simulating the behavior of the SHA model in Figure 7.8. For details on the model and the associated reliability analysis, readers could refer to our publication in [42].

$$\boldsymbol{SHA = (Q, E, X, A, A_c, H, F, P, q_0, x_0, P_0)}$$

$Q = \{1,2\}$	System states: 1-OK, 2-Failure
$E = \{e_{11}, e_{12}\}$	e_{11} - One damage shock
	e_{12} -One fatal shock
$X = \{x, \sum W\}$	Degradation x and accumulated damage $\sum W$
$A = \{A_{11}, A_{12}\}$	$A_{11} = \{1, e_{11}, G_{11}, R_{11}, 1\}$, $A_{12} = \{1, e_{12}, G_{12}, R_{12}, 2\}$
	$G_{11}: \{t = rand(h_{11})\}, R_{11}: \sum W \coloneqq \sum W + \alpha Y_i, h_{11} \coloneqq 0$
	$G_{12}: \{t = rand(h_{12})\}, R_{12}: x \coloneqq D, \sum W \coloneqq H, h_{12} \coloneqq 0$
$A_c = (A_{c1}, A_{c2})$	See the Figure
$H = (h_{11}, h_{12})$	h_{11}: damage shock arrival time,
	h_{12}: fatal shock arrival time.
$F = (F_{11}, F_{12})$	$F_{11} : exp(P_1\lambda(t)), F_{12} : exp(P_2\lambda(t))$

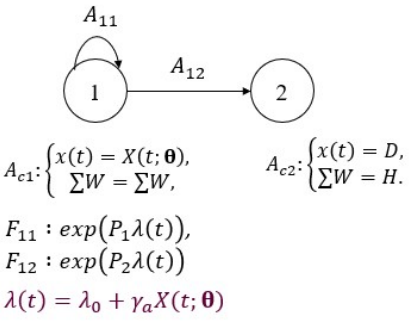


Figure 7.8: SHA modeling for the sliding spool.

7.3 Efficient analysis of dependent failure processes based on stochastic hybrid systems

In this section, we aim at the second research problem identified in Sect. 7.1. To improve the efficiency of the reliability analysis, we introduce a semi-analytical framework based on a special type of stochastic hybrid automation, i.e., the stochastic hybrid system (SHS). The SHS model is briefly reviewed in Sect. 7.3.1. Then, the dependent degradation-shock processes are modeled as an SHS in Sect. 7.3.2. In Sects. 7.3.3 and 7.3.4, we present how to assess the reliability efficiently based on the SHS model. The work in this section is based on one of our publication [43].

7.3.1 SHS model

The state space of a SHS model is a combination of discrete and continuous states. Let us denote the discrete states by $q(t), q(t) \in Q$, where Q is a finite set containing all the possible discrete modes of the system. The continuous states are denoted by $x(t), x(t) \in \mathbb{R}^l$. A SHS model is defined based on the following assumptions [57, 56, 55]:

(1) The evolution of the continuous states is governed by a set of SDEs:

$$dx(t) = f(q(t), x(t))\, dt + g(q(t), x(t))\, dw_t, \tag{7.1}$$

where $w_t : \mathbb{R}^+ \to \mathbb{R}^k$ is a k-dimensional Wiener process; $f : Q \times \mathbb{R}^l \to \mathbb{R}^l$ and $g : Q \times \mathbb{R}^l \to \mathbb{R}^{l \times k}$, respectively.

(2) At any time t, if the system is in state $(q(t), x(t))$, it undergoes a transition with a rate $\lambda_{ij}(q(t), x(t)) : Q \times \mathbb{R}^l \to \mathbb{R}^+$, $i, j \in Q$. That is, the probability that the system undergoes a transition from state i to state j within the interval $[t, t + \Delta t)$ is:

$$\lambda_{ij}(q(t), x(t))\, \Delta t + o(\Delta t), \tag{7.2}$$

(3) Whenever the system undergoes a state transition from state i to state j , it instantaneously applies the map $\phi_{ij}(q(t), x(t))$ to the current values of $q(t)$ and $x(t)$, so that their values are reset:

$$((q(t), x(t))) = \phi_{ij}\left((q(t^-), x(t^-))\right), \tag{7.3}$$

where the notation $a(t^-)$ represents the left-hand limit of the function a at time t. Figure 7.9 summarizes the assumptions and depicts the state transition and evolution of the SHS.

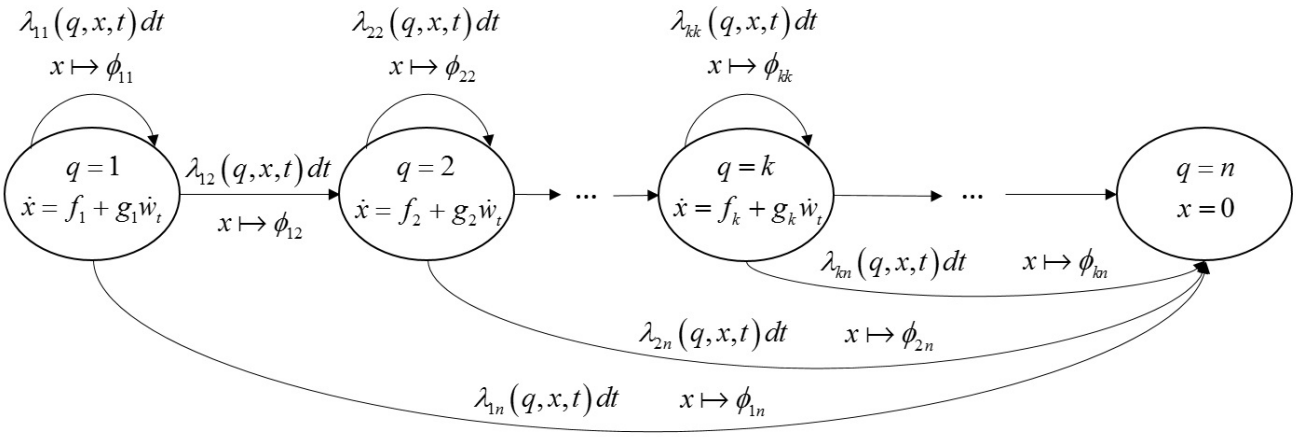


Figure 7.9: State transition diagram for the SHS model.

7.3.2 SHS formulism for dependent failure processes

The modeling framework for dependent failure processes involves three elements, i.e., a model for the degradation process, a model for the shock process and a model for the dependency between the two processes. The following assumptions are made in order to model a dependent failure process in the framework of SHS:

Assumption (1): The degradation processes are characterized by $x(t) = (x_1(t), x_2(t), ..., x_l(t)) \in \mathbb{R}^l$. The elements in $x(t)$, $x_i(t), 1 \leq i \leq l$, are performance parameters for the degradation processes and are independent from one another. Soft failure occurs whenever $\exists i \in \{1, 2, \cdots, l\}, x_i(t) > H_i$, where H_i is the failure threshold for the performance parameter $x_i(t)$.

Assumption (2): The system has n potential health states, i.e., $q(t) \in Q$ where $q(t)$ is a discrete-state variable that quantifies the system's health state at time t , and $Q = \{1, 2, ..., n\}$ is a set containing all the possible system states. When $q(t) = n$, a hard failure occurs.

Assumption (3): Transitions between system health states are triggered by the arrival of random shocks with the transition rate $\lambda_{ij}(q(t), x(t))$, $i, j \in Q$, where the probability that the system jumps from state i to state j in the interval $[t, t + \Delta t)$ is given by Eq. (7.2).

Assumption (4): Between the transitions, the degradation of $x(t)$ is characterized by the SDEs in Eq. (7.1) for $q(t) = 1, 2, \cdots, n-1$. When $q(t)$ takes different values, the form of $f(\cdot)$ and $g(\cdot)$ can be changed to reflect the dependency behavior. When $q(t) = n$, which indicates that the system fails due to hard failure, we impose that $x(t) = 0$.

Assumption (5): An arrival random shock resets the current values of $q(t)$ and $x(t)$, using the reset map defined in Eq. (7.3).

Assumption (6): System failure is caused by both soft and hard failures, whichever occurs first.

Given a dependent failure process, the following steps show how to model it in the framework of SHS:

Step 1: Modeling degradation. In this step, the performance parameters $x(t)$ are identified to characterize the degradation processes. For the performance parameters, the SDEs in Eq. (7.1) are developed to describe their degradation, considering both deterministic and stochastic characteristics. The deterministic characteristics are often described based on the physical knowledge on the degradation processes (e.g., using the Physics-of-Failure (PoF) models [97]), while the stochastic characteristics are modeled by a Wiener process, as shown in Eq. (7.1).

Step 2: Modeling random shocks. In SHS, random shocks are considered as transitions among the system health states. The transition rates, $\lambda_{ij}(q(t), x(t))$, $i, j \in Q$, need to be determined based on historical data or expert judgments.

Step 3: Modeling dependencies. Finally, the dependencies between the degradation processes and random shocks need to be considered. The dependencies can be modeled in various ways in SHS. For instance, by resetting the values for $x(t)$, the reset map in Eq. (7.3) can capture the influence of the random shock on the degradation process. Further, the functions f, g and even λ itself, as shown in Figure 7.9, are dependent on the current values of $x(t)$ and $q(t)$, which provides a versatile way to model the dependencies.

Note that in order to make sure the developed SHS model is solvable in case of truncations techniques [57] are needed, for example Case 3 in this paper, the $f_i, g_i, \lambda_{ij}, \phi_{ij}, i, j \in Q$ in the SHS model have to be polynomial functions of $x(t)$.

7.3.3 Conditional moments estimation

In this section, we derive the conditional expectations for the continuous state variables, i.e., $E\left[x_j^p(t) | q(t) = i\right]$, $p \in \mathbb{N}, i \in Q, j = 1, 2, \cdots, l$, where $x_j(t)$ represent the j th element of $x(t)$. The conditional expectations will be used in the next section for reliability analysis. Let us define a test function to be

$$\psi_i^{(m)}(q, x) = \begin{cases} x^m, q = i, \\ 0, q \neq i. \end{cases} \tag{7.4}$$

where $m := (m_1, m_2, ..., m_l), m \in \mathbb{N}^l$, and $x^m := x_1^{m_1} x_2^{m_2} \cdots x_l^{m_l}$, and let the m -order conditional moment of the continuous state x be

$$\begin{aligned} \mu_i^{(m)}(t) &:= E[\psi_i^{(m)}(q, x)] \\ &= E[x^m(t) \mid q(t) = i] \cdot Pr(q(t) = i). \end{aligned} \tag{7.5}$$

For a general test function $\psi(q(t), x(t))$, $\psi : Q \times \mathbb{R}^l \to \mathbb{R}$, which is twice continuously differentiable with respect to x , the evolution of its expected value is governed by Dynkin's formula [57]:

$$\frac{dE[\psi(q(t), x(t))]}{dt} = E[(L\psi)(q(t), x(t))] \tag{7.6}$$

where $(L\psi)(q,x)$ is the extended generator of SHS and $\forall (q,x) \in Q \times \mathbb{R}^l$, $(L\psi)(q,x)$ is given by

$$\begin{aligned}(L\psi)(q,x) &:= \frac{\partial \psi(q,x)}{\partial x} f(q,x) \\ &+ \frac{1}{2} trace\left(\frac{\partial^2 \psi(q,x)}{\partial x^2} g(q,x)\, g(q,x)'\right) \\ &+ \sum_{i,j \in Q} \lambda_{ij}(q,x)\left(\psi(\phi_{ij}(q,x)) - \psi(q,x)\right),\end{aligned} \tag{7.7}$$

where $\partial\psi/\partial x$ and $\partial^2\psi/\partial x^2$ denote the gradient and Hessian matrix of $\psi(q,x)$ with respect to x , respectively; $trace(A)$ is the trace of the matrix A , i.e., the sum of elements on its main diagonal.

Substituting Eq. (7.4) into (7.6), we get a group of differential equations with respect to $\mu_i^{(m)}(t), i \in Q, m \in \mathbb{N}^l$:

$$d\mu_i^{(m)}(t) = E\left[L\left(\psi_i^{(m)}\right)(q(t), x(t))\right] \cdot dt. \tag{7.8}$$

The evolution of $\mu_i^{(m)}(t)$ can be depicted by solving Eq. (7.8). The conditional moments can, then, be obtained by assigning proper values for m : if we let $m = (0, 0, ..., 0)$, we have

$$\mu_i^{(0,0,\ldots,0)}(t) = \Pr\{q(t) = i\}, i \in Q. \tag{7.9}$$

If we let

$$m = [m_1, m_2, \cdots, m_l] : \begin{cases} m_j = p, & \text{if } j = k, k \in \{1, 2, \cdots, l\}, \\ m_j = 0, & \text{if } j \neq k, \end{cases}$$

where m_j denotes the j th element in m and p is a natural number, we have

$$\mu_i^{(m)}(t) = E\left[x_k^p(t) \,|q(t) = i\right] \cdot \Pr\{q(t) = i\}, i \in Q. \tag{7.10}$$

The conditional expectations, $E\left[x_j^p(t) \,|q(t) = i\right]$, $p \in \mathbb{N}, i \in Q, j = 1, 2, \cdots, l$, can, then, be calculated by combining Eqs. (7.9) and (7.10).

7.3.4 Reliability analysis

From Assumption 6, system reliability can be expressed as:

$$R(t) = \Pr(q(t) < n, x_1(t) < H_1, x_2(t) < H_2, \cdots, x_l(t) < H_l).$$

From the law of total probability, we have

$$\begin{aligned} R(t) &= \Pr\left(q\left(t\right) < n, x_1\left(t\right) < H_1, x_2\left(t\right) < H_2, \cdots, x_l\left(t\right) < H_l\right) \\ &= \sum_{i=1}^{n-1} Pr(q(t) = i) \cdot Pr(x_1(t) < H_1, x_2(t) < H_2, \cdots, x_l(t) < H_l \mid q(t) = i). \end{aligned} \tag{7.11}$$

Since we assume that the degradation processes are independent from one another, Eq. (7.11) becomes

$$R\left(t\right) = \sum_{i=1}^{n-1} \left(\prod_{j=1}^{l} \Pr\left(x_j\left(t\right) < H_j \left| q\left(t\right) = i\right.\right) \right) \cdot \Pr\left(q\left(t\right) = i\right) \tag{7.12}$$

In Eq. (7.12), $\Pr\left(q\left(t\right) = i\right)$ can be calculated by (7.9), $Pr(x_j(t) < H_j \mid q(t) = i)$, $i = 1, 2, \cdots, n-1$, $j = 1, 2, \cdots, l$ can, instead, be approximated using the First Order Second Moment (FOSM) method [165], since we have the conditional moments for $x_j\left(t\right)$. Let $\mu_{x_j|q=i}\left(t\right)$ and $\sigma_{x_j|q=i}\left(t\right)$ denote the expected value and standard deviation of the random variable $x_j\left(t\right)$ conditioned on $q = i$, respectively. Then, $\mu_{x_j|q=i}\left(t\right)$ and $\sigma_{x_j|q=i}\left(t\right)$ can be calculated by

$$\begin{aligned} \hat{\mu}_{x_j|q=i}(t) &= E[x_j(t)|q(t) = i] = \frac{\mu_i^{(m^*,j)}(t)}{Pr(q(t) = i)} = \frac{\mu_i^{m^{(*,j)}}(t)}{\mu_i^{(0,0,\ldots,0)}(t)} \\ \hat{\sigma}_{x_j|q=i}(t) &= \sqrt{E\left[x_j(t)^2|q(t) = i\right] - \left(E[x_j(t)|q(t) = i]\right)^2} \\ &= \sqrt{\frac{\mu_i^{(m^{**},j)}(t)}{\mu_i^{(0,0,\ldots,0)}(t)} - \left(\frac{\mu_i^{(m^*,j)}(t)}{\mu_i^{(0,0,\ldots,0)}(t)}\right)^2} \end{aligned} \tag{7.13}$$

where $m^{*,j}$ and $m^{**,j}$ are given by

$$\begin{aligned} m^{*,j} &= [m_1, m_2, \cdots, m_l] : m_k = 1, \text{ if } k = j; m_k = 0, \text{ if } k \neq j, \\ m^{**,j} &= [m_1, m_2, \cdots, m_l] : m_k = 2, \text{ if } k = j; m_k = 0, \text{ if } k \neq j. \end{aligned}$$

Based on FOSM, $\Pr\left(x_j\left(t\right) < H_j \left| q\left(t\right) = i\right.\right)$ can be approximated by

$$\Pr\left(x_j\left(t\right) < H_j \left| q\left(t\right) = i\right.\right) \approx \Phi\left(\frac{H_j - \hat{\mu}_{x_j|q=i}\left(t\right)}{\hat{\sigma}_{x_j|q=i}\left(t\right)}\right). \tag{7.14}$$

Substituting Eq. (7.14) into (7.12), the reliability of the system is approximated by

$$R\left(t\right) \approx R_e\left(t\right) = \sum_{i=1}^{n-1} \mu_i^{(0,0,\ldots,0)}\left(t\right) \cdot \left(\prod_{j=1}^{l} \Phi\left(\frac{H_j - \hat{\mu}_{x_j|q=i}\left(t\right)}{\hat{\sigma}_{x_j|q=i}\left(t\right)}\right) \right),$$

where $\hat{\mu}_{x_j|q=i}\left(t\right)$, $\hat{\sigma}_{x_j|q=i}\left(t\right)$ are calculated by Eq. (7.13).

The accuracy of the approximation by FOSM relies on the normality assumption: the random variables $x_j\left(t\right) \left| q\left(t\right) = \right.$

$i, i \in 1, 2, ..., n-1, j = 1, 2, \cdots, l$ are normally distributed with mean value $\mu_{x_j|q=i}(t)$ and standard deviation $\sigma_{x_j|q=i}(t)$. In practice, the assumption does not always hold. Therefore, we also present an estimation method for the lower bound of the system reliability, using Markov inequality.

According to Markov inequality [112], if X is a nonnegative random variable and $a > 0$, then

$$\Pr(X \geq a) \leq \frac{E(X)}{a}. \tag{7.15}$$

Using Eq. (7.15), we obtain

$$\Pr(x_j(t) \geq H_j | q = i) \leq \frac{E(x_j(t) | q = i)}{H_j}, j \in \{1, 2, ..., l\}, i \in \{1, 2, ..., n-1\}. \tag{7.16}$$

From Eqs. (7.12) and (7.16), the lower bound of system reliability can, then, be derived:

$$R(t) \geq R_l(t) = \sum_{i=1}^{n-1} \mu_i^{(0,0,...,0)}(t) \cdot \Pi_{j=1}^{l} \left[1 - \frac{\mu_i^{(m^{*,j})}(t)}{H_j}\right] \tag{7.17}$$

where the parameters have the same meaning as in Eq. (7.13).

We test the developed reliability framework on four widely-used benchmark case studies from literature, and compare its performance with Monte Carlo simulations. The results show that the developed framework can accurately estimate the reliability while significantly reducing the computational time. Details of the application could be found in [43].

7.4 Analyzing common cause failure with the SHS-based framework

In this section, we extend the SHS-based framework to model and analyze system failure behaviors that involve common cause failures with dependent degrading components. The extended modeling framework is presented in Sect. 7.4.1. An efficient reliability assessment method is presented in Sect. 7.4.2. In Sect. 7.4.3, the developed approaches are applied on a real-world case study. The works in this section were previously published as a journal paper [44]. For more information, readers could refer to the original publication in [44].

7.4.1 SHS-based modelling framework for CCFs of degrading components

Let us consider a generic system with l degrading components. It is assumed that the CCFs are caused by random shock processes. Two types of shocks are distinguished:

- lethal shock, whose arrival causes simultaneous failures of all the components in the associated CCCG;

- non-lethal shock, whose arrival causes an additional damage to the degradation process of all the components and a failure is caused when the cumulative damage reaches its threshold.

Examples of lethal shocks include product design flaws, equipment miscalibration, catastrophic environmental conditions, etc. Examples of non-lethal shocks include unusual temperature or vibration, debris in a shared fluid, etc. It should be noted that when a non-lethal shock arrives, a component fails with a probability which depends on the current state of its degradation process.

The following assumptions are made to model the system in the framework of SHS modelling:

(1) The degradation processes of the l degrading components of the system are described in terms of the vector $x(t) = (x_1(t), x_2(t), \ldots, x_l(t)) \in \mathbb{R}^l$, where $x_j(t)$, $1 \leq j \leq l$, denote the degradation level of the jth component and $x_j(t) \geq H_j$ indicates the failure of the jth component.

(2) The degradation of $x(t)$ can be modeled using the stochastic differential equations, as described in the SHS framework in Sect. 7.3.

(3) The system suffers from n lethal shocks. The i th lethal shock causes the failure of all the components in the associated CCCG, denoted by $CCCG_i$, $i = 1, 2, \cdots, n$.

(4) The system is also subject to non-lethal shocks, which contribute cumulatively to the degradation processes of all the components.

(5) System reliability is modeled by its structure function.

An SHS model is illustrated in Figure 7.10, where the continuous variable $x(t)$ describes the degradation process. The discrete variable $q(t) \in \{0, 1, 2, \cdots, n\}$ describes the lethal shocks: $q(t) = 0$ indicates that no lethal shock arrived, whereas $q(t) = i, i = 1, 2, \cdots, n$ indicates that the i th lethal shock arrived and caused a CCF. Transitions between system states are triggered by the arrival of random shocks (both non-lethal or lethal), with transition rates $\lambda_{ij}(q(t), x(t))$, $i, j \in Q$, where $i = j$ indicates a non-lethal shock and $i \neq j$ indicates a lethal shock. In practice, the transition rates $\lambda_{ij}, i, j \in Q$, need to be estimated based on historical shock data or expert judgments. Between successive transitions, the degradation of $x(t)$ is governed by the stochastic differential equations. In practice, these SDEs can be determined based on physics-of-failures [**?**].The variability in the degradation processes is usually described using the Wiener process. The reset map $\phi_{ij}, i, j \in Q$ quantifies the effect of the shock process: when a non-lethal shock arrives, the degradation levels of all components are reset by an increment (constant or random); when the kth lethal shock arrives, the degradation levels of the components in $CCCG_k$ are reset to their thresholds.

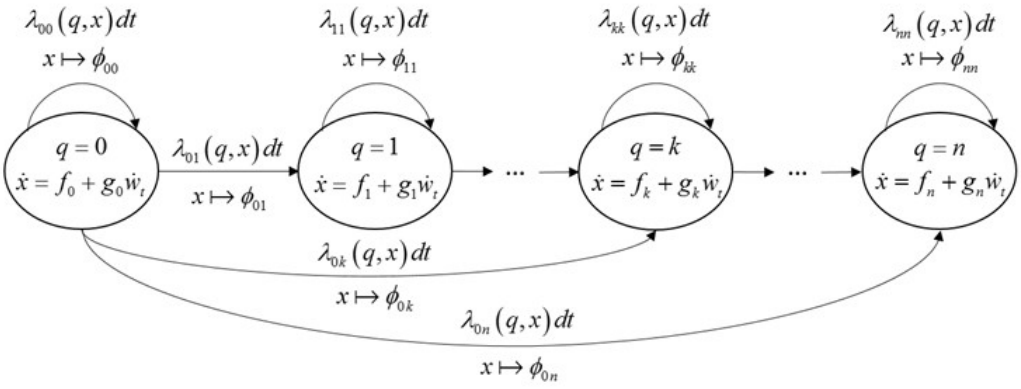


Figure 7.10: SHS model for CCF.

7.4.2 System reliability analysis

It is assumed, from assumption (5) in Sect. 7.4.1, that the system reliability is modeled by its structure function. Let us suppose that the system structure function is

$$Y_S = F(Y_1, Y_2, ..., Y_l), \tag{7.18}$$

where Y_S and $Y_1, Y_2, ..., Y_l$ are Boolean variables representing system and component states, respectively, where Y_S and $Y_i = 1$ indicate working states. Given the structure function, the system reliability can be expressed explicitly as a function of the components reliabilities, if the components are independent:

$$\begin{aligned} R_S(t) &= \Pr\{F(Y_1, Y_2, \cdots, Y_l) = 1\} \\ &= G(R_1(t), R_2(t), ..., R_l(t)), \end{aligned} \tag{7.19}$$

where $R_j(t), j = 1, 2, ..., l$, denotes the reliability of the *j*th component, and $G(\cdot), G : [0,1]^l \to [0,1]$, is determined according to $F(\cdot)$.

In the SHS model, the components are not independent due to the existence of CCFs. However, if we condition on the system state $q = k, k \in 1, 2, ..., n$, the component failures become conditional independent. Therefore, the system reliability can be calculated from the law of total probability:

$$R_S(t) = \sum_{i=0}^{n} \Pr(q(t) = i) \cdot G\left(R_{1|q=i}(t), R_{2|q=i}(t), ..., R_{l|q=i}(t)\right), \tag{7.20}$$

where $R_{j|q=i}(t)$ denotes the conditional reliability of the *j*th component on condition that $q = i$.

In Eq. (7.20), $\Pr(q(t) = i)$ and $R_{j|q=i}(t)$ can be estimated using the SHS-based framework presented in Sect.

7.3. Hence, the reliability of the system is approximated by

$$R_S(t) \approx \sum_{i=0}^{n} \mu_i^{(0,0,\ldots,0)}(t) \cdot G\left(\Phi\left(\frac{H_1 - \hat{\mu}_{x_1|q=i}(t)}{\hat{\sigma}_{x_1|q=i}(t)}\right), \Phi\left(\frac{H_2 - \hat{\mu}_{x_2|q=i}(t)}{\hat{\sigma}_{x_2|q=i}(t)}\right), ..., \Phi\left(\frac{H_l - \hat{\mu}_{x_l|q=i}(t)}{\hat{\sigma}_{x_l|q=i}(t)}\right)\right), \quad (7.21)$$

where $\hat{\mu}_{x_j|q=i}(t), \hat{\sigma}_{x_j|q=i}(t), j = 1, 2, ..., l, i \in Q$, are estimated through the SHS framework.

The accuracy of the approximation by the FOSM method relies on the normality assumption: the random variables $x_j(t)|q(t) = i, i \in 0, 1, ..., n, j = 1, 2, \cdots, l$, are normally distributed with mean value $\mu_{x_j|q=i}(t)$ and standard deviation $\sigma_{x_j|q=i}(t)$.

7.4.3 Application

System descriptions

In this section, we show how to apply the SHS-based framework to model the CCF using a real-world example of an Auxiliary Feedwater Pump (AFP) in a Nuclear Power Plant (NPP). In an auxiliary feedwater system, AFPs may fail due to internal flood from three main water sources, i.e. Service Water (SW), Circulating Water (CW) and Fire Protection Water (FPW) [138]. Piping rupture in any of the three water systems can cause an internal flood that destroys the AFP. According to [95], the observed three most common modes of piping rupture include random rupture, seismic-induced rupture and tornado induced rupture, the first of which is modeled as degradation-induced failure in this paper. To protect AFPs from the internal flood, a flood barrier is built in the safeguards alley where AFPs are located. However, if the barrier breaks, mostly due to degradation or intensive earthquakes, AFPs will also fail when there is an internal flood. The above described failure mechanism for AFP failures due to internal flood is presented as a Fault Tree (FT) in Figure , where A failure, B failure, C failure and D failure represent “SW piping rupture”, “CW piping rupture”, “FPW piping rupture” and “flood barrier break”, respectively.

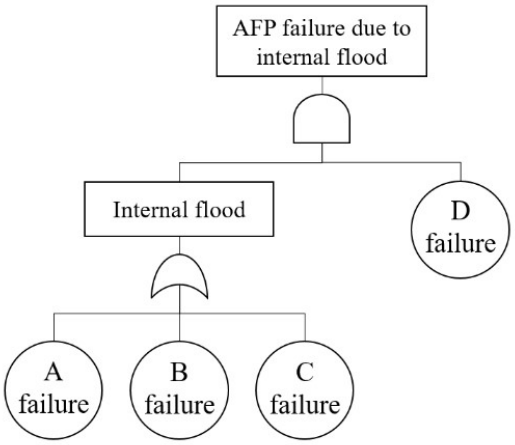


Figure 7.11: Fault tree for ”AFP failure due to internal flood”.

(1) The components are subject to stochastic degradation processes:

$$dx_j(t) = \alpha_j dt + \beta_j dw_t, j = A, B, C, D \tag{7.22}$$

where $x_j(t)$ denotes the degradation measure of component j , $w_t \in \mathbb{R}$ is a standard Wiener process, α_j, β_j are degradation constants. Also, we assume that $x_j(0) = 0, \forall j \in \{A, B, C, D\}$. A component fails when its degradation reaches the corresponding failure threshold.

(2) The system is subject to tornadoes (lethal shock 1) and earthquakes (lethal shock 2), which follow HPPs with intensities λ_t and λ_e , respectively; when a tornado occurs, components A, B, C fail due to the CCF, i.e., $CCCG_1 = \{A, B, C\}$; when an earthquake occurs, all components fail due to the CCF, i.e., $CCCG_2 = \{A, B, C, D\}$. It should be noted that in this paper, we assume that two lethal shocks cannot occur simultaneously. This assumption is valid because the probability that two or more lethal shocks occur within a very short time interval, i.e. $(t, t + \Delta t), \Delta t \to 0$, is usually very low in practice, since the individual lethal shocks are usually rare events. Similar assumptions are often adopted in reliability modelling. For example, in Markov reliability models, it is often assumed that only one component can fail in a very short time interval.

(3) The system is also subject to non-lethal shocks, which follow an HPP with intensity λ_{nl} ; when a non-lethal shock occurs, a degradation increment d_j occurs to component j , which follows a normal distribution $d_j \sim N\left(\mu_{d_j}, \sigma^2_{d_j}\right), j = A, B, C, D$.

SHS model and reliability analysis

The SHS for the system is shown in Figure 7.12. The system has three states $q(t) \in \{0, 1, 2\}$: when $q(t) = 0$, the system is in the normal operation state, which means no lethal shock occurs before *t*. Components degrade according to Eq. (7.22); when a non-lethal shock occurs, the degradation processes of the four components are reset adding an increment $d_j \sim N\left(\mu_{d_j}, \sigma^2_{d_j}\right), j = A, B, C, D$, respectively; when $q(t) = 1$, components contained in $CCCG_1$, i.e. components *A*, *B*, *C*, fail simultaneously and their degradation levels are reset to their respective thresholds, while component *D* degrades according to ; when $q(t) = 2$, components contained in $CCCG_2$, i.e. components *A*, *B*, *C*, *D*, fail and their degradation levels are reset to their thresholds, respectively. For the SHS model in this case, $l = 4, n = 2$.

Reliability analysis can, then, be conducted based on the developed framework. Details of the analysis can be found in our publication [44]. Figure 7.13 briefly summarized the results from the analysis. It can be seen that compared to the benchmark model (the BFR model, which is a widely-used CCF model from literature), the developed framework can more accurately capture the actual failure behavior, as it is able to model the dependent degradation behaviors of the components. Also, the developed framework can greatly reduce the compuational costs of the reliability analysis, as it does not rely on Monte Carlo simulation for the analysis.

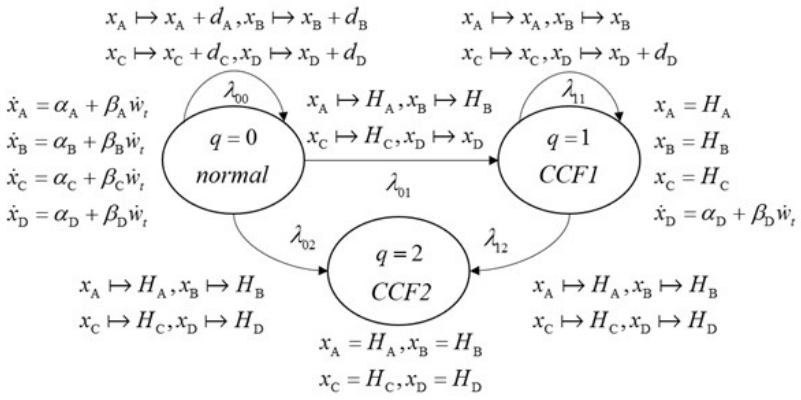


Figure 7.12: State-transition diagram of the SHS for the AFP system.

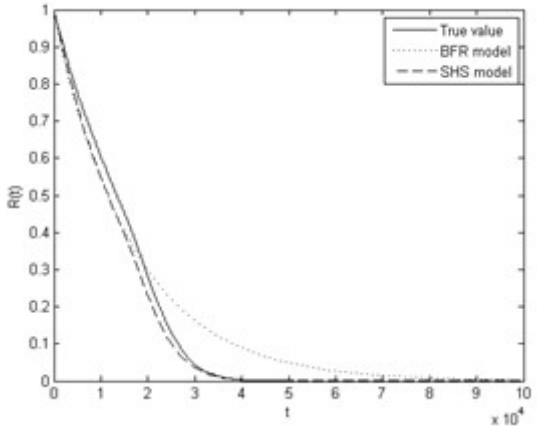


Figure 7.13: Results of SHS and BFR for the AFP system.

7.5 Summary of major contributions

In this chapter, we presented our major works related to modeling and analysis of dependent degradation-shock processes. Driven by the two research questions identified in Sect. 7.1, the main contributions of our works can be summarized as follows:

1. We developed a generic modeling framework, based on stochastic hybrid automaton, for dependent degradation-shock processes. We tested the framework on a few widely-used benchmark examples from literature, and also on a real-world case study. The results show that our model allows modeling a variety of dependent degradation-shock behaviors through a unified framework. In this way, the developed framework has the potential to greatly reduce the amount of effort in modeling dependent failure behaviors.

2. We developed a semi-analytical reliability assessment framework for dependent degradation-shock processes. Based on a mathematical framework of SHS, the developed method allows estimating reliability through solving a series of differential equations, whose numerical solutions are very efficient. We tested the perfor-

mance of our methods on four widely-used benchmark examples from literature. The results showed that the developed methods can greatly improve the computational efficiency of reliability assessment of dependent degradation-shock processes.

3. We developed an SHS-based model for systems subject to common cause failures and dependent degrading components. Efficient reliability assessment algorithms are also derived. Compared to existing models for CCF, our work allows considering the dependent failure behaviors of the components in the system, which is often neglected in the traditional CCF models. Hence, the developed model can more accurately describe the system failure behaviors.

Chapter 8

QUANTIFYING EPISTEMIC UNCERTAINTY AND ITS IMPACT ON RISK AND RELIABILITY MODELS

This chapter summarizes some of my representative research results in research axis 4. The focus of this research axis is to develop methods to quantify the impact of epistemic uncertainty on risk and reliability assessment. In Sect. 10.1, we briefly review the related literature and define the research problems of this research axis. Some representative results are briefly introduced in Sects. 10.2 and 10.3: in Sect. 10.2, we present a classification-based framework to quantify the impact of epistemic uncertainty on a probability risk assessment; in Sect. 10.3, we present a hierarchical framework for evaluating the degree of trustworthiness (a measure of epistemic uncertainty) in PRA, and integrate the degree of trustworthiness in the result of risk quantification. Finally, in Sect. 10.5, we summarize the major contributions achieved in this research axis.

8.1 Research questions

In risk/reliability analysis, numerical metrics are calculated based on models that describe the the occurrence and evolution of failure and its behavior. It is well known that epistemic uncertainty (EU) exists in these models, referring to the uncertainty that results from incomplete/insufficient knowledge and/or approximations of the processes involved in the failure/accidents [159]. Here, we follow the operational perspective of the US Nuclear Regulatory Commission to classify EU into completeness uncertainty, model structural uncertainty and parametric uncertainty, as shown in Figure 8.1 [107], and review the existing researches on quantifying EU in risk and reliability.

Completeness uncertainty results from the fact that the PRA might be incomplete and fail to consider some

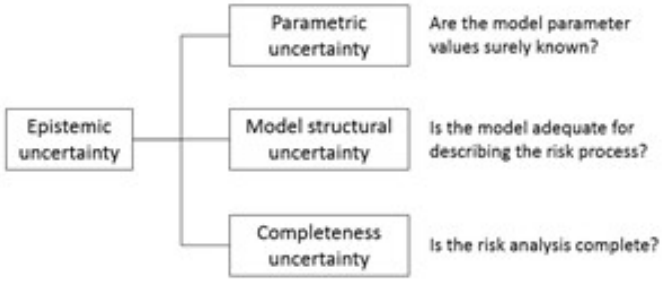


Figure 8.1: A classification of EU in risk and reliability analysis.

important risk contributors [107]. This might result in underestimation of risk [158]. Terms and concepts used in the literature in relation to completeness uncertainty include ignorance, surprising events, black swan [13], etc. Unforeseen accident scenarios caused by completeness uncertainty have been discussed extensively in the risk analysis literature. For example, Kaplan and Garrick [77] proposed a Bayesian framework to consider unforeseen scenarios, in which an artificially constructed scenario called “others” is added in the analysis to reflect the lack of completeness. The total risk is calculated based on the law of total probability and can be updated using Bayesian methods when new observation data become available [77]. Kazemi and Mosleh [79] applied a similar method to investigate the impact of surprising events on credit risks. Bjerga et al. 19 discussed the exact meaning of completeness uncertainty and proposed a practical approach for handling it in risk assessment. They concluded that completeness uncertainty can be treated as model uncertainty. Most of these works, however, are conceptual: operational guidelines to support their practical implementation are not provided.

Model structural uncertainty (also referred to as model uncertainty in some papers) arises from the way the PRA model accounts for the features of the processes involved 8. Because of model structural uncertainty, systematic errors might be introduced into the predicted risk indexes [158]. For example, static PRA models like event tree fail to capture risk indexes that are time-dependent due to various degradation mechanisms [152]. Mosleh and Droguett reviewed the common approaches used for characterizing model structural uncertainty [38]. Among them, the alternate hypotheses approach and the adjustment factor approach are listed as two most widely applied ones [170]. The alternate hypotheses approach develops an overall PRA model by probabilistically combining several alternate models, each of which is developed under alternate assumptions of the model structures [36]. The probabilistic combination is done by Bayesian model averaging, where the weights of the alternate models are determined from experimental data or expert judgements that measure closeness of the models to reality [37]. In the adjustment factor approach, an adjustment factor is added to or multiplied by the prediction result of a reference PRA model to describe the influence of model structural uncertainty [119]. Mosleh and Apostolakis [102] applied the adjustment factor approach to evaluate the influence of model uncertainty on a seismic risk assessment based on experts’ judgements.

Parametric uncertainty relates to the estimated values of parameters of the PRA model [107]. Usually, it results

in a “level-two” uncertainty analysis setting where outer loop simulations sample realizations of variables subject to epistemic uncertainty (denoted by E), while for each outer loop simulation, inner loop simulations are conducted to sample from the variables subject to aleatory uncertainty, conditioned on the realizations of E (see [110] for details). Various mathematical frameworks have been developed for quantifying and propagating parametric uncertainty, e.g., probability theory, evidence theory, possibility theory, probability box, and interval analysis [54]. For example, Hao et al. [51] applied the probability-based framework to consider the parametric uncertainty in a risk assessment of a water inrush accident in a karst tunnel. Xie et al. [140] used evidence theory to describe the parametric uncertainty in a PRA model of a pressure vessel subject to corrosion and developed a kriging model-based adaptive sampling method for effective risk assessment.

Although a substantial amount of works have been done, as reviewed above, two issues still remain to be addressed:

1. most existing methods for EU quantification only apply to a specific type of EU (either completeness uncertainty, model structural uncertainty or parametric uncertainty). A unified framework that is capable to consider the three sources of EU collectively is lacking.

2. directly applying these frameworks in practice, however, is sometimes difficult, as it is not easy to determine the required information (e.g., probability distributions representative of the actual state of EU) from an actual risk/reliability assessment. Operational methods and guidelines, are, thus, needed for quantifying EU in practice.

In this chapter, we focus on these two research questions. Section 8.2 focuses the first research question by proposing a maturity model for epistemic uncertainty management. Sections 10.2 and 10.3 discuss the second question: in Sect. 10.2, a machine learning-based approach is proposed for the numeration of epistemic uncertainty, while in Sect. 10.3, a hierarchical framework with evaluation guidelines is proposed.

8.2 Maturity model for epistemic uncertainty management

In this section, we present an unified framework to quantify the three aspects of EU, by developing a maturity model for epistemic uncertainty management. The maturity model is presented in Sect. 10.3.1. Section 10.3.2 shows detailed definitions of the maturity levels. Section 10.3.3 shows how to use the developed model for quantifying epistemic uncertainty. The work in this section was published in our journal paper [160]. More details on the model and its application could be found in the original paper.

8.2.1 The model

Various types of EU might affect the PRA model building process (Figure 8.2). If not properly managed, the EU could impact the results of the PRA and the decisions made based on these results. For example, the first steps of a PRA lead to identifying scenarios that need to be analyzed. Insufficient/incomplete knowledge in these steps would lead to completeness uncertainty. Therefore, the resulting PRA model would not cover all possible scenarios and possibly underestimate the risk. Once the scenarios are identified, models of the evolution of the scenarios are built to compute the risk index for different possible consequences. Model structural uncertainty might be introduced in this part: the model might not fully describe the real physical process, and, as a result, systematic errors in the risk indexes might occur. Finally, in the calculation, parametric uncertainty related to the estimation of the model parameters, might lead to inaccurate risk index values and, as a result, affect the decisions made based on these.

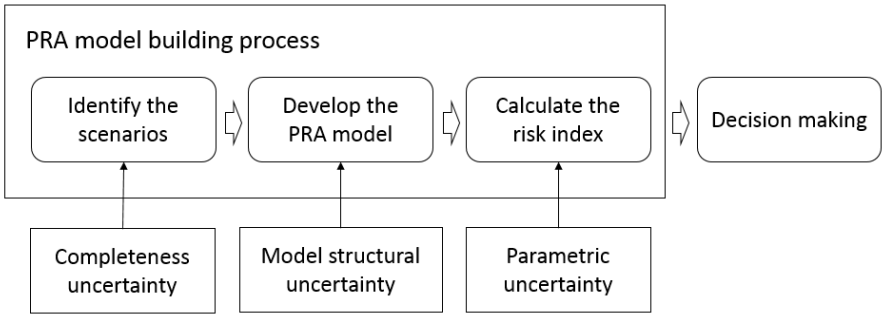


Figure 8.2: The EU that affects a PRA.

For properly informed decision making, the EU in the PRA needs to be managed. In this paper, we define Epistemic Uncertainty Management (EUM) capability as the ability to identify, characterize and control the epistemic uncertainty in a PRA model. Here, we use the term “epistemic uncertainty” in a broad sense, *i.e.*, it covers completeness uncertainty, model structural uncertainty and parametric uncertainty. EUM must allow evaluating the EU in the PRA model and for this a maturity model for EUM (MM-EUM) is developed in this paper.

Similar to the work on the capability maturity model for software development processes , MM-EUM is a framework to capture the key elements which enable EUM in PRA. The MM-EUM represents an evolutionary improvement from *ad hoc* EU management to strengthened EU management capability in PRA. This is expected to yield more transparent and trustworthy PRA results, and better support for risk-informed decision making.

The structure of the MM-EUM comprises three elements (Figure 8.3): maturity levels, activities and goals. Five maturity levels are defined to describe different degrees of EUM in PRA. The five levels are distinguished based on the severity of the potential impact of the EU on the PRA results. A detailed definition of the maturity levels can be found in Sect. 10.3.2. For each maturity level, several activities that help to generate the corresponding level of maturity are identified (Sect. 10.3.3). Each activity is associated with one or several goals. If all of the goals at a given maturity level are fulfilled, the PRA reaches such maturity level.

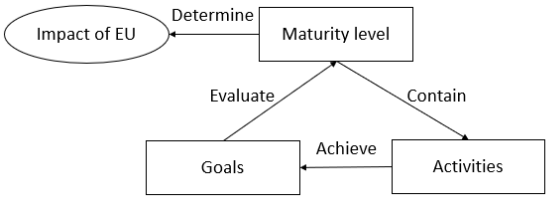


Figure 8.3: The structure of MM-EUM.

8.2.2 Maturity levels

Let M_{EUM} denote the maturity of EUM for a PRA. Based on the severity of the influence of EU on the PRA results, we define five levels of M_{EUM}:

- Initial $(M_{EUM} = 1)$: The PRA is conducted without considering the influence of EU. The sources of EU and their influence on the result of the PRA are unknown and unmanaged.
- Uncontrolled $(M_{EUM} = 2)$: The PRA is conducted with an epistemic uncertainty analysis (covering completeness, model structural and parametric uncertainty). The potential impact of EU is known to the decision maker, but no measures have been taken to reduce it.
- Complete $(M_{EUM} = 3)$: Effective measures have been taken to control the completeness uncertainty (reduce its impact to a desired level). As a result, the PRA is complete: the critical risk contributors that might severely affect the results of the PRA have all been considered in the analysis, given the current knowledge and the degree of accuracy required.
- Adequate $(M_{EUM} = 4)$: Effective measures have been taken to control the model structural uncertainty. As a result, the developed PRA model is capable to adequately capture the characteristics of the process involved in the risk assessment, given the current knowledge and the degree of accuracy required.
- Accurate $(M_{EUM} = 5)$: Effective measures have been taken to control the parametric uncertainty. As a result, the parameters in the risk assessment model are estimated to the required level of accuracy.

As shown in Figure 8.4, the five maturity levels defined above characterize a cumulative improvement process of the EUM in PRA. The improvement process starts from the Initial level At this level, no analysis has been conducted to identify the possible sources of EU in the PRA. The PRA is conducted without considering the possible influence of EU.

At the Uncontrolled level $(M_{EUM} = 2)$, the sources of EU in the PRA process have been identified. Through the analysis, the impact of completeness uncertainty, model structural uncertainty and parametric uncertainty are known and quantified. EU has been characterized and propagated into the PRA result using proper mathematical theories. However, the EU is not controlled: no measures are implemented to contain and reduce the existing EU.

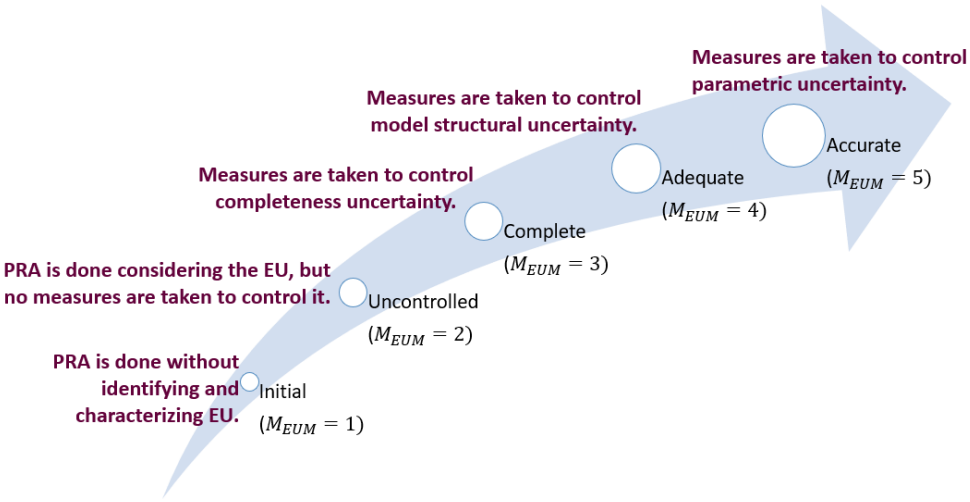


Figure 8.4: Continuous improvement process of the maturity levels.As shown in Figure3, the five maturity levels defined above characterize a cumulative improvement process

At the Complete level $(M_{EUM} = 3)$, not only the achievements of the Uncontrolled level are obtained, the completeness uncertainty is actively controlled as well, through the activities defined in Sect. 10.3.3. Reaching this level indicates that the EU management is capable to control the completeness uncertainty, so that all the important risk contributors have been considered in the PRA model (to the current knowledge level).

At the Adequate level $(M_{EUM} = 4)$, besides the achievements of the previous levels, the model structural uncertainty is actively controlled through the activities defined in Sect. 10.3.3. Arriving at this level means that the PRA model is adequate in terms of its capability to account for the actual physical processes involved. Therefore, no significant systematic errors are expected to exist in the results of the PRA.

At the Accurate level $(M_{EUM} = 5)$, in addition to the achievements of the previous levels, the parametric uncertainty is controlled through the activities defined in Sect. 10.3.3. If a PRA reaches the Accurate level, the EU is properly controlled and one can be confident that the calculated risk index reflects all the available knowledge on the risk as well as the uncertainties.

In our framework, the Uncontrolled level already requires that a “complete” epistemic uncertainty analysis is done, considering completeness, model structural and parametric uncertainties. The difference between the Uncontrolled level and subsequent three levels is that, in the Uncontrolled level, the decision maker only knows how uncertain he/she is due to the impact of EU, but does not take any proactive measures; while starting from the Complete level, proactive measures are taken to reduce the impacts of EU. It should be noted that the orderings of the maturity levels are defined based on the severity of the potential impact of EU for a particular maturity level. For example, the Complete level $(M_{EUM} = 3)$ is considered as less mature than the Adequate and Accurate levels, as the potential impact of completeness uncertainty is more severe than that of model and parametric uncertainty: we should make sure first that we are modelling the correct risk contributors, before considering if we had chosen an

appropriate model (model uncertainty) for analysis and if the parameter values are accurately estimated (parametric uncertainty). The maturity model defined in Figure 8.4 also provides a sequential process to guide the activities of improvement for reducing epistemic uncertainty: the requirements of a lower maturity level should be satisfied first, before considering the requirements of a higher maturity level.

8.2.3 Activities and their goals

With the exception of the Initial level, each maturity level can be achieved by effectively implementing several key activities that support it. To verify if the key activities are implemented successfully, several goals are defined for each key activity: if all the goals for the key activities at a given maturity level i are successfully fulfilled, the corresponding maturity level i is reached, *i.e.*, $M_{EUM} = i$. In Table 8.1, we present an example of the activities and their associated goals that support the maturity levels 2. Similar Tables are also defined for the other maturity levels in our published paper [160].

Table 8.1: Key activities and associated goals for the Uncontrolled level ($M_{EUM} = 2$).

Key activities	Goals
Document the PRA	• The PRA needs to be documented in a well-organized report. • The report should contain the necessary information for identifying sources of EU, including completeness uncertainty, model structural uncertainty and parametric uncertainty.
Identify the sources of EU	• Potential sources of EU need to be identified through an analysis conducted by qualified experts. • The analysis needs to cover completeness uncertainty, model structural uncertainty and parametric uncertainty. • The results of the analysis need to be confirmed by peer reviews from independent experts.
Analyze the impacts of EU	• The impact of the EU on the calculated risk indexes needs to be analyzed by qualified experts. • The analysis should cover completeness uncertainty, model structural uncertainty and parametric uncertainty. • Decision making considers both the calculated risk indexes and the EU in the PRA.

The activities and goals defined in the Tables can be used to evaluate the maturity level in EUM:

- if all the goals of the activities for a given maturity level i are fulfilled, the corresponding maturity level is considered as being reached, *i.e.*, $M_{EUM} = i$;
- otherwise, we have $i - 1 < M_{EUM} < i$. The precise value is determined by experts based on the degree to which the goals are satisfied.

Besides, the activities and goals can also be used to plan the efforts needed to control the EU in the PRA.

Suppose that the current maturity level is i : To improve the EUM capability of the PRA, one needs to focus on the activities and unsatisfied goals at maturity level $i+1$.

The developed maturity model has been applied in a real-world case study in our original publication [160]. We did not put details here due to page limits. Interested readers could directly refer to [160].

8.3 A classification-based framework for trustworthiness assessment of quantitative risk analysis

In this section, we use trustworthiness of risk assessment to represent the degree of epistemic uncertainty on its result and develop a classification-based method for the assessment of the trustworthiness of Quantitative Risk Analysis (QRA). The assessment framework is presented in Sect. 8.3.2. Then, in Sect. 8.3.2, we present a machine learning algorithm for evaluating the trustworthiness of QRA, based on Naive Bayes classifier. An application of the developed frameworks is presented in Sect. 8.3.3. The work in this section was originally published as a journal paper [150]. More details could be found in the published paper.

8.3.1 Assessment framework

Let T represent the trustworthiness of QRA. We take a proactive perspective on trustworthiness assessment and assume that T is determined by the quality of the QRA process. According to Rae et al. [116], a typical QRA process involves eight sub-processes, as shown in Figure 8.5. To ensure the quality of a QRA process, all the eight sub-processes should be conducted with high quality [116]. A framework for trustworthiness assessment is, then, developed in Figure 8.6 by considering the quality requirements on the eight sub-processes in Figure 8.5.

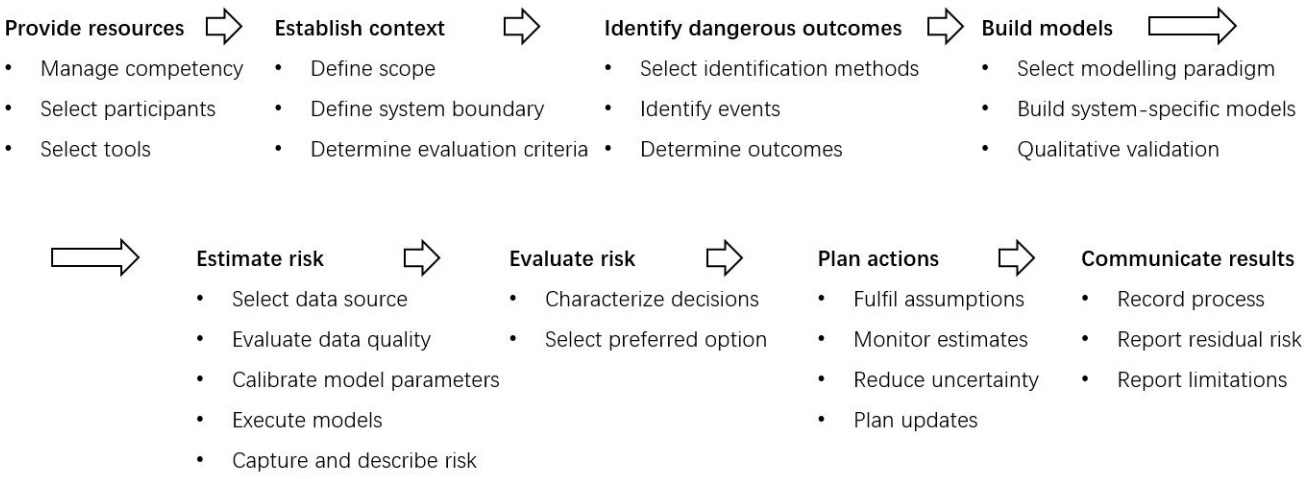


Figure 8.5: A typical QRA process [116]

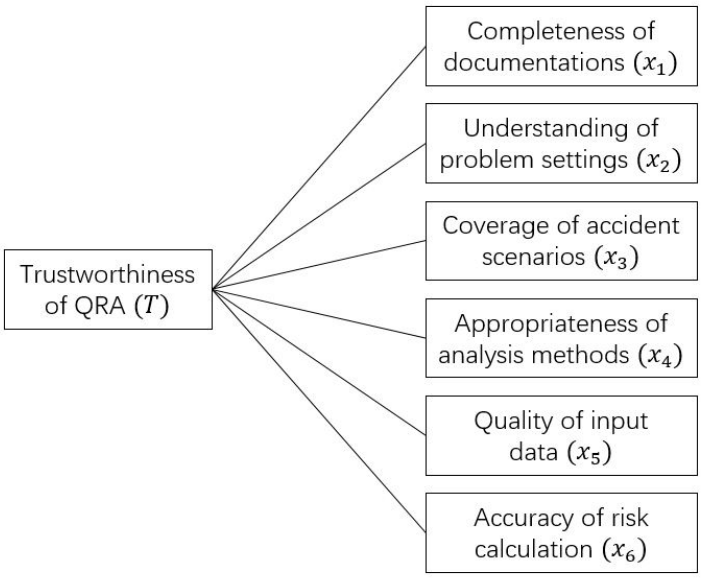


Figure 8.6: Trustworthiness assessment framework

In Figure 8.6, the trustworthiness of QRA is characterized in terms of six criteria, i.e., completeness of documentations (x_1), understanding of problem settings (x_2), coverage of accident scenarios (x_3), appropriateness of analysis methods (x_4), quality of input data (x_5), accuracy of risk calculation (x_6), which reflect the quality requirements on the QRA process. Each criterion is evaluated into three grades, i.e., problematic ($x_i = 0$), acceptable ($x_i = 1$) and satisfactory ($x_i = 2$), $i = 1, 2, \cdots, 6$, based on a set of predefined scaling rules. In Table 8.3, we illustrate the scaling rule for x_1. The scaling rules for the other variables can be found in our paper [150]. Three discrete levels of T, i.e., $T \in \{0, 1, 2\}$, are considered in this paper. The levels are distinguished in Table 8.2 based on their reliability, which concerns the repeatability of the risk analysis [12] and validity, which concerns whether the risk analysis addresses the "right problem" [12]. The problem of trustworthiness assessment is, then, formulated as a classification problem: given the states of the six criteria $x_1, x_2, \cdots, x_6$, determine an appropriate category for the trustworthiness T. It should be noted that both the assessment framework in Figure 8.6 and the scaling rules are constructed for illustrative purposes. They are defied in a general form that allows them to be adapted for capturing the problem-specific features in practical applications.

8.3.2 Trustworthiness assessment based on Naive Bayes classifier

For a given QRA, the values of x_1-x_6 can be decided by doing an evaluation based on the defined grading rules. How to use this information to assess the trustworthiness T still remains a problem. In this section, we present a classification-based framework that allows extracting expert knowledge on the mapping from x_1-x_6 to T for trustworthiness assessment.

Table 8.2: Three levels for T

Levels of trustworthiness	Descriptions
$T = 0$: Unreliable	• The result of the QRA is unrepeatable. • No further judgements can be made on the trustworthiness of the QRA. • Such QRA should not be used to support any decision making.
$T = 1$: Reliable but invalid	• The result of QRA is repeatable but • some critical hazards are not identified and analyzed by the QRA or • some important risks (and their uncertainties) are not accurately quantified by the QRA. • Such QRA can be used to support decision making, but not for safety-critical decisions.
$T = 2$: Reliable and valid	• The result of the QRA is repeatable and • all critical hazards are identified and analyzed by the QRA; • all important risks (and their uncertainties) are accurately quantified by the QRA. • Such QRA can be used to support critical decision making.

Table 8.3: Scaling rules for x_1

Levels	Descriptions
$x_1 = 0$	Some the following elements are missing in the documentations: • scopes and objectives of the QRA; • descriptions of the system under investigation and related references; • accounts of the adopted analysis methods; • presentation of source data needed for the analysis; • report of the analysis results.
$x_1 = 1$	At least one of the following flaws present in the documentations: • descriptions of scopes and objectives are incomplete or ambiguous; • descriptions of the system under investigation are unclear; • no sufficient references on the system under investigation are given; • descriptions of the adopted methods are unclear; • presentations of the results are incomplete (e.g., no uncertainty is considered) or ambiguous.
$x_1 = 2$	The documentation of the QRA process contains sufficient information for its repetition: • the documentation contains all the necessary parts; • no flows in level $x_1 = 1$ present.

Basics of naive Bayes classifier

Let us define $\mathbf{x} = [x_1, x_2, \cdots, x_n] \in \mathbb{X}$ to be the input feature vector of the classification problem, where $\mathbb{X}$ is the feature space. A NBC is a function f_{NBC} that maps input feature vectors $\mathbf{x} \in \mathbb{X}$ to output class labels $T \in \{0, 1, \cdots, C\}$ [3]. Usually, the feature vector also takes discrete values, so that we have $x_i \in \{0, 1, \cdots, n_i\}, i = 1, 2, \cdots, n$. Given a feature vector $\mathbf{x}$, a NBC classifies it into the class with the maximum posterior probability [3]:

$$T = \arg\max_{T} Pr(T \mid \mathbf{x}). \tag{8.1}$$

The posterior probability in (8.1) is calculated using Bayes rule [3]:

$$Pr(T \mid \mathbf{x}) = \frac{Pr(\mathbf{x}, T)}{Pr(\mathbf{x})} = \frac{Pr(\mathbf{x} \mid T)Pr(T)}{\sum_{T=0}^{C} Pr(\mathbf{x} \mid T)Pr(T)}. \tag{8.2}$$

If we further assume that the elements $x_i, i = 1, 2, \cdots, n$ of the input feature vector $\mathbf{x}$ are independent, the

nominator of (8.2) becomes:

$$Pr(\mathbf{x} \mid T)Pr(T) = Pr(T)\prod_{i=1}^{n} Pr(x_i \mid T). \tag{8.3}$$

Note that the denominator in (8.2) is the same for all possible values of T. Therefore, (8.1) can be simplified:

$$T = \arg\max_{T} Pr(T)\prod_{i=1}^{n} Pr(x_i \mid T). \tag{8.4}$$

In order to apply the NBC, the $Pr(T)$ and $Pr(x_i|T)$ in (8.4) should be estimated from training data. Training data are a set of samples whose correct classes are already known. Suppose we have $N_{training}$ training data, denoted by $(\mathbf{x}^{(q)}, T^{(q)}), q = 1, 2, \cdots, N_{training}$. Then, the required probabilities are estimated by:

$$Pr\left(T = k\right) = \frac{\sum_{q=1}^{N_{training}} \mathbb{1}\left(T^{(q)} = k\right)}{N_{training}}, \tag{8.5}$$

$$Pr\left(x_i = j \mid T = k\right) = \frac{\sum_{q=1}^{N_{training}} \mathbb{1}\left(x_i^{(q)} = j, T^{(q)} = k\right)}{\sum_{q=1}^{N_{training}} \mathbb{1}\left(T^{(q)} = k\right)}, \tag{8.6}$$

where $\mathbb{1}(\cdot)$ is the indicator function and $i = 1, 2, \cdots, n, j = 0, 1, \cdots, n_i, k = 0, 1, \cdots, C$.

There is one potential problem for (8.5) and (8.6). Suppose that due to statistical variations, for some specific values of j and k, we have $\sum_{q=1}^{N_{training}} \mathbb{1}\left(x_i^{(q)} = j, T^{(q)} = k\right) = 0$. In this case, $Pr\left(x_i = j \mid T = k\right) = 0$, which, according to (8.3), results in $Pr(\mathbf{x}|T) = 0$, regardless of the posterior probabilities for other features. Misclassification often happens in such situations. To avoid such a problem, a technique called Laplacian correction is often applied when estimating $Pr\left(T = k\right)$ and $Pr\left(x_i = j \mid T = k\right)$ [3]:

$$Pr\left(T = k\right) = \frac{\sum_{q=1}^{N_{training}} \mathbb{1}\left(T^{(q)} = k\right) + \gamma}{N_{training} + (C+1)\cdot\gamma}, \tag{8.7}$$

$$Pr\left(x_i = j \mid T = k\right) = \frac{\sum_{q=1}^{N_{training}} \mathbb{1}\left(x_i^{(q)} = j, T^{(q)} = k\right) + \gamma}{\sum_{q=1}^{N_{training}} \mathbb{1}\left(T^{(q)} = k\right) + (n_i+1)\cdot\gamma}, \tag{8.8}$$

where $\gamma \in (0, 1]$ is an adjustment factor introduced to compensate for the possible zero probabilities; $C + 1$ and $n_i + 1$ are the number of possible values for T and x_i, respectively.

Trustworthiness assessment based on NBC

In this section, we apply the NBC to develop a classifier for the trustworthiness assessment problem in Figure 8.6. In this case, we have six features, i.e., $\mathbf{x} = [x_1, x_2, \cdots, x_6]^T$. Each feature has three discrete levels, i.e., $x_i \in \{0, 1, 2\}, i = 1, 2, \cdots, 6$. Hence, $\mathbb{X} = \{0, 1, 2\} \times \cdots \times \{0, 1, 2\} = \{0, 1, 2\}^6$. The trustworthiness also takes three

values, i.e., $T \in \{0, 1, 2\}$.

Training data collection

Since $\mathbb{X} = \{0, 1, 2\}^6$, the feature vector **x** can take $3^6 = 729$ different values. A fraction of them, denoted by $\mathbf{x}^{(q)}, q = 1, 2, \cdots, N_{training}$, are selected as training samples. The trustworthiness of these training samples, denoted by $T^{(q)}, q = 1, 2, \cdots, N_{training}$, are evaluated by experts, based on the descriptions in Table 8.2. The training data are, then, used to construct the NBC and once constructed, the NBC is exploited to replace the expert for the assessment of trustworthiness.

Since the NBC learns the expert's evaluation rationale from the training data, it is essential that the training data are a reasonable representation of the whole feature space. On the other hand, we want to reduce the number of training data as much as possible, since collecting training data is often expensive and time-consuming. For this, in this paper, we use an experiment design technique, i.e., the row-exchange algorithm in Matlab R2015b, to design the training data collection scheme. The response model in the row-exchange algorithm is assumed to be a linear model and the resulted D-optimal design matrix is used for the collection of training data. This approximates an orthogonal design on the $\mathbf{x}^{(q)}, q = 1, 2, \cdots, N_{training}$, where the collected training data are equally distributed and can equally "represent" the entire space of $\mathbb{X}$.

Another issue that needs to be considered when designing the training data collection scheme is the sample size $N_{training}$. Apparently, a large value of $N_{training}$ would enhance the performance of the developed classifier in terms of its accuracy. On the other hand, large values of $N_{training}$ also create more difficulties in collecting the data (experts easily get impatient when asked to judge too many scenarios). Hence, a trade-off needs to be made in determining the value of $N_{training}$.

Construction of the classifier

The procedures for constructing the NBC is summarized in Figure 8.7. In the preparation phase, the sample size of the training set and the training data collection scheme are determined using the methods discussed previously. The training data $(\mathbf{x}^{(q)}, T^{(q)}), q = 1, 2, \cdots, N_{training}$ are, then, collected by expert judgements following the scaling rules in Table 8.3 and the others. In the training phase, the NBC is constructed by estimating $Pr(T)$ and $Pr(x_i|T)$ from the training data, using (8.7) and (8.8), respectively. In the evaluation phase, the constructed NBC is applied to replace the role of the experts and determine the trustworthiness of a new QRA. By reviewing the related documents, the value for the feature vector **x** of the QRA is determined first, based on the scaling rules defined in Table 8.3 and the others. Its trustworthiness is, then, determined based on the constructed NBC using (8.4).

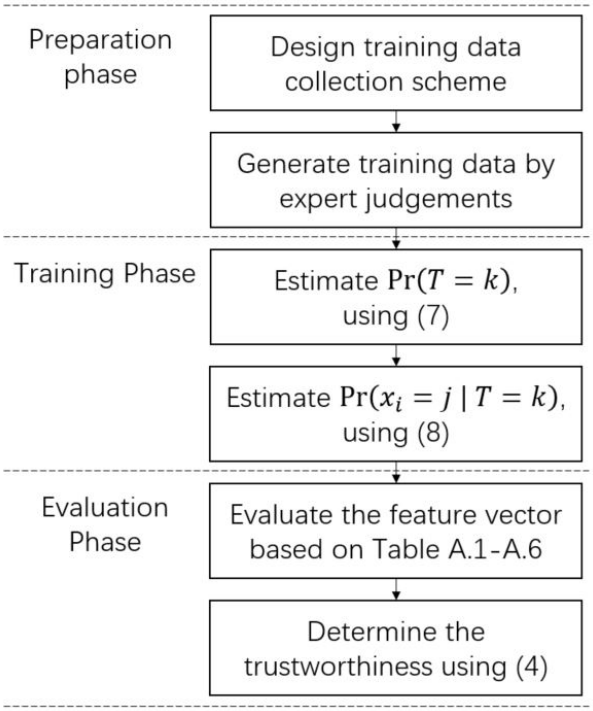


Figure 8.7: NBC construction procedure for QRA trustworthiness assessment

8.3.3 Application

In this section, we show how to apply the developed framework to assess the trustworthiness of a real-world methanol plant, wherein the associated individual and social risks are assessed by a systematic QRA process, in terms of risk contours and F-N curve, respectively [5]. The training data used for the construction of the NBC are generated by asking an expert to assess the trustworthiness of a set of artificially generated "pseudo" QRAs. The quality criteria of the methanol QRA is evaluated by reviewing its final report, which is available online from [5]. The main results are summarized as follows.

Training data collection scheme

In this step, we design the training data collection scheme. From our previous discussion, we can see that $N_{training} = 54$ can, in general, yield good classification accuracy. Therefore, we choose $N_{training} = 54$. The row-exchange algorithm in Matlab R2015b is used to design the training data collection scheme. The resulting $\mathbf{x}^{(q)}, q = 1, 2, \cdots, N_{training}$ are listed in Table 8.4. It can be verified that the training data collection scheme in Table 8.4 is an orthogonal design. The values of $\mathbf{x}^{(q)}, q = 1, 2, \cdots, N_{training}$ correspond to the levels of the quality criteria in Table 8.3 and the others.

Training data collection

Each row in Table 8.4 represents a pseudo QRA, characterized by specific quality criteria. An expert is asked to assess the trustworthiness for these pseudo QRAs, for generating the training data. Take the first row in Table 8.4 as an example. To generate the training data, the expert is asked the following question: if the quality of a QRA process is as depicted in Table 8.5, which level of trustworthiness in Table 8.2 do you think the QRA has? Table 8.5 is generated by relating the values of $\mathbf{x}^{(q)}, q = 1, 2, \cdots, N_{training}$ to the corresponding quality criteria in Table 8.3 and the others. The procedures are repeated for the other rows in Table 8.4. The training data generated by the expert are also listed in Table 8.4.

Table 8.4: Training data

Runs	$x_1^{(q)}$	$x_2^{(q)}$	$x_3^{(q)}$	$x_4^{(q)}$	$x_5^{(q)}$	$x_6^{(q)}$	$T^{(q)}$	Runs	$x_1^{(q)}$	$x_2^{(q)}$	$x_3^{(q)}$	$x_4^{(q)}$	$x_5^{(q)}$	$x_6^{(q)}$	$T^{(q)}$
1	0	0	0	0	1	0	0	28	1	1	1	2	1	1	1
2	0	0	0	0	2	2	0	29	1	1	2	1	0	1	1
3	0	0	1	2	1	1	0	30	1	1	2	2	1	2	1
4	0	0	1	2	2	2	0	31	1	2	0	1	2	0	1
5	0	0	2	1	0	1	0	32	1	2	0	2	0	1	1
6	0	0	2	1	1	1	0	33	1	2	1	0	1	0	1
7	0	1	0	1	0	0	0	34	1	2	1	0	1	2	1
8	0	1	1	0	2	1	0	35	1	2	2	0	2	0	1
9	0	1	1	2	0	0	0	36	1	2	2	2	0	2	2
10	0	1	2	0	0	0	0	37	2	0	0	0	0	2	0
11	0	1	2	0	1	2	0	38	2	0	0	2	1	0	0
12	0	1	2	2	2	0	0	39	2	0	1	1	0	2	1
13	0	2	0	1	1	1	0	40	2	0	1	2	2	0	1
14	0	2	0	1	1	2	0	41	2	0	2	0	0	0	0
15	0	2	0	2	2	0	0	42	2	0	2	2	2	1	2
16	0	2	1	0	2	1	0	43	2	1	0	0	0	1	0
17	0	2	1	2	0	2	0	44	2	1	0	1	2	2	1
18	0	2	2	1	0	2	0	45	2	1	0	2	1	0	1
19	1	0	0	0	1	2	0	46	2	1	1	1	1	1	1
20	1	0	0	2	0	1	0	47	2	1	1	1	2	2	1
21	1	0	1	1	0	2	1	48	2	1	2	0	1	2	1
22	1	0	1	1	1	0	1	49	2	2	0	0	0	1	0
23	1	0	2	0	2	1	1	50	2	2	1	0	2	1	1
24	1	0	2	1	2	0	1	51	2	2	1	1	0	0	1
25	1	1	0	1	2	1	1	52	2	2	2	1	1	0	1
26	1	1	0	2	2	2	1	53	2	2	2	2	1	1	2
27	1	1	1	0	0	0	0	54	2	2	2	2	2	2	2

Classifier construction

The training data are used to construct the NBC, following the procedures in Figure 8.7. The accuracy of the constructed classifier is evaluated by the correct classification rate and we have $CR = 0.944$. Therefore, the constructed NBC can be used to represent the expert judgements and provide reasonable assessment of the trustworthiness of QRA.

Table 8.5: Quality of the first pseudo QRA

Criteria	Level
Completeness of documentation	Some the following elements are missing in the documentations: • scopes and objectives of the QRA; • descriptions of the system under investigation and related references; • accounts of the adopted analysis methods; • presentation of source data needed for the analysis; • report of the analysis results.
Understanding of problem settings	The analysts are unaware of the problem settings of the QRA due to the presence of all the following flaws: • the purposes of the QRA are not clearly understood; • the systems of interests are not well defined; • the resources constraints (e.g., time, computational resources, etc) are not clearly defined.
Coverage of accident scenarios	Some critical accident scenarios are highly likely to be missed by the identification process: • the coverage of the identified accident scenarios is not validated; • the validation shows that some critical accident scenarios might be missing.
Appropriateness of analysis methods	• The features of the selected analysis method satisfy the requirements of the problem and • successful applications in similar problems can justify the choice of the method.
Quality of input data	• There is no sufficient statistical data and the input data is purely based on expert judgements; • epistemic uncertainty in the expert-generated input data is not considered.
Accuracy of risk calculation	• Only errors from the calculation process itself (e.g., the accuracy of Monte Carlo simulations) might exist and • the uncertainties caused by the errors are properly modeled.

The constructed NBC can also help to explain the expert's behavior in assessing the trustworthiness. For example, from the training results, we notice that $Pr\left(x_1 = 0 \mid T = 0\right) = 0.6882$, $Pr\left(x_1 = 0 \mid T = 1\right) = 0.0041$, $Pr\left(x_1 = 2 \mid T = 0\right) = 0.0233$. From Bayes theorem,

$$
\begin{aligned}
Pr\left(T = 0 \mid x_1 = 0\right) &= \frac{Pr\left(x_1 = 0 \mid T = 0\right) \cdot Pr(T = 0)}{Pr(x_1 = 0)} \\
&= \frac{Pr\left(x_1 = 0 \mid T = 0\right) \cdot Pr(T = 0)}{\sum_{i=1}^{3} Pr\left(x_1 = 0 \mid T = i\right) \cdot Pr(T = i)} \qquad (8.9) \\
&= 0.9896 \qquad (8.10)
\end{aligned}
$$

That is, if x_1 equals to zero, the expert tends to judge the QRA as unreliable. This is a natural result, since x_1 denotes the completeness of documentations. If the QRA process is not well-documented, it is unlikely to be repeatable: therefore, the associated QRA is unreliable according to the criteria in Table 8.2.

Table 8.6: A comparison to existing methods

Methods	Correct classification rate
Classification-based method	$CR = 0.944$
Conformance-based method	$CR = 0.130$, for $n_{th} = 0$
	$CR = 0.315$, for $n_{th} = 1$
	$CR = 0.463$, for $n_{th} = 2$
	$CR = 0.556$, for $n_{th} = 3$
	$CR = 0.500$, for $n_{th} = 4$
	$CR = 0.500$, for $n_{th} = 5$
	$CR = 0.482$, for $n_{th} = 6$

Comparison to existing methods

In traditional proactive trustworthiness assessment methods, e.g., [6], expert knowledge is elicited to develop a simple conformance/non-conformance-based framework that relates the quality criteria to the trustworthiness of the QRA. That is, the conclusion of whether the QRA is trustworthy or not is made by comparing the number of the conformed quality criteria to a predefined threshold value n_{th}. In this paper, we assume that a quality criterion i is conformed when $x_i = 2$. Table 8.6 shows a comparison between the classification-based framework and the conformance/non-conformance-based framework, using the training data in Table 8.4. It can be seen that in general, the existing conformance/non-conformance-based framework cannot accurately model the complex expert knowledge expressed in the empirical data in Table 8.4. The developed method, on the other hand, is capable of capturing the complex behavior of expert judgement in assessing the trustworthiness of the QRA.

Trustworthiness assessment

To assess the trustworthiness of the methanol QRA using the developed NBC, its six quality criteria are first evaluated based on the QRA report [5] and following the scaling rules in Table 8.3 and the others. For example, the scaling rule for completeness of documentation (x_1) is listed in Table 8.3. In general, the methanol QRA report contains sufficient information on the scope and objective of the analysis, the system under investigation and the adopted analysis methods. However, according to Table 8.3, the presentation of the analysis results is incomplete, since no accounts of uncertainty are given in the report. Therefore, we have $x_1 = 1$. The other elements can be evaluated in a similar way. Hence, we have $\mathbf{x} = [1, 2, 1, 0, 0, 2]$. By running the NBC with the input feature vector $\mathbf{x} = [1, 2, 1, 0, 0, 2]$, we can calculate the posterior probabilities from (8.4), as shown in Figure 8.8. We can conclude that $T = 1$ for the QRA of the methanol plant, which means, according to Table 8.2, that the QRA of the methanol plant is reliable but invalid. Such a QRA can be used to support decision making, but not for safety-critical decisions.

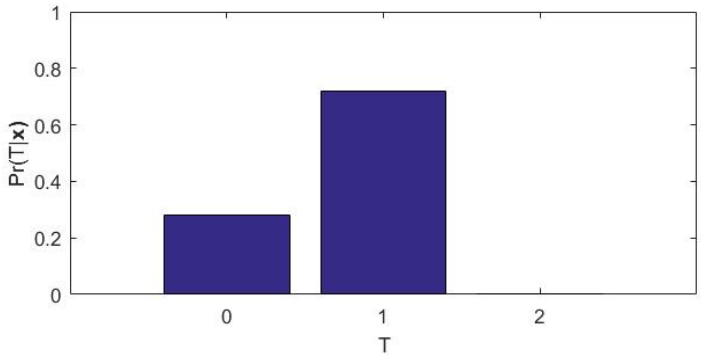


Figure 8.8: Posterior probabilities for each value of T

8.4 Multi-hazards risk aggregation considering trustworthiness

In this paper, we develop a new method for Multi-Hazards Risk Aggregation (MHRA) considering trustworthiness of the risk assessment. A hierarchical framework is first developed for evaluating the trustworthiness of the risk assessment (Sect. 8.4.1). The trustworthiness is calculated using a weighted average of the leaf attributes, in which the weights are calculated using the Dempster Shafer Theory-Analytical Hierarchy Process (DST-AHP) (Sects. 8.4.2 and 8.4.3). Risk aggregation is, then, performed by a "weighted posterior" method, considering the level of trustworthiness (Sect. 8.4.4). An application to the risk aggregation of two hazard groups in Nuclear Power Plants (NPP) is illustrated in Sect. 8.4.5. More details of the work in this section can be found in our publication [18].

8.4.1 Evaluation of the level of trustworthiness

In this section, a bottom-up method for evaluating the level of trustworthiness is developed. Five levels of trustworthiness are defined with their corresponding settings:

- Strongly untrustworthy ($T = 1$): represents the minimum level of trustworthiness and, therefore, the decision maker has the lowest confidence in the result of the PRA. The analysis is made based on weak knowledge and/or nonrealistic analysis, leading to an estimated value that might be far from the real one. Further analysis and justifications need to be implemented on the risk analysis to enhance its trustworthiness. Otherwise, the risk assessment is not considered representative and one should not rely on its results to support any kind of decision making.

- Untrustworthy($T = 2$): represents a low level of trustworthiness and, therefore, the decision maker has low confidence in the results of the PRA. At this level, the analysis is made based on relatively weak knowledge and/or nonrealistic analysis, leading to unrealistically estimated risk values. Further analysis and justifications need to be implemented on the risk analysis to enhance its trustworthiness. The decision maker can use the results with caution and only as a support for decision making.

- Moderately trustworthy ($T = 3$): represents a moderate level of trustworthiness and, therefore, the decision maker has an acceptable level of confidence in the results of the PRA. The analysis is made based on relatively moderate knowledge and/or relatively realistic analysis. The decision maker can rely cautiously on the model output to make the decision.
- Trustworthy ($T = 4$): represents a high level of trustworthiness and, therefore, the decision maker has quite high confidence in the results of the PRA. The analysis is made on a relatively high level of knowledge and realistic analysis. The decision maker can rely confidently on the models output to make decisions.
- Highly trustworthy ($T = 5$): represents the maximum level of trustworthiness. At this level, the PRA model outputs accurately predict the risk index with a proper characterization of parametric uncertainty. The decision maker can rely on the models output to support decision making involving severe consequences, e.g., loss of human lives.

In practice, the trustworthiness of risk assessment might be between two of the five levels defined above: for example, $T = 2.6$ means that the level of trustworthiness is between untrustworthy and moderately trustworthy.

Based on a thorough literature review, we develop a hierarchical framework that breaks down the contributing factors of T into a tree structure, comprising of lower-level attributes and sub-attributes that are more tangible to be assessed. The hierarchical framework is shown in Figure 8.9. For detailed descriptions of the attributes and sub-attributes, please refer to our paper in [18].

In this paper, the level of trustworthiness of risk assessment is evaluated using a weighted average of the "leaf" attributes in Figure 8.9.

$$T = \sum_{i}^{n} W_i \cdot A_i \tag{8.11}$$

where W_i is the weight of the leaf attribute that measures its relative contribution to the trustworthiness of risk assessment; A_i is the trustworthiness score for the *i*-th leaf attribute, evaluated based on the scoring guidelines presented in the Appendixes of [18]; n is the number of the leaf attributes (in Figure 8.9, we have $n = 27$). The weights $W_i$are determined based on Dempster Shafer-Analytical Hierarchy Process (DST-AHP) [35], as discussed in the next section.

8.4.2 Dempster Shafer Theory - Analytical Hierarchy Process (DST-AHP) for trustworthiness attributes weight evaluation

The weights of the different attributes in Figure 8.9 can be determined using the AHP method to compare their relative importance with respect to the trustworthiness of risk assessment [122]. AHP is used because it can decrease the complexity of the comparison process, as it allows comparing only two criteria at a time, rather than comparing all the criteria simultaneously, which could be very difficult in complex problems. It should be noted that

since there are no alternatives to be compared in this framework, pairwise comparison matrixes of AHP are only used for deriving the attributes (criteria) weights.

To consider the fact that experts are subjective, not fully reliable and might have conflicting viewpoints, as well as considering the incomplete knowledge of the experts, Dempster-Shafer-Analytical Hierarchy Process (DST-AHP) is used. This allows combining multiple sources of uncertain, fuzzy and highly conflicting pieces of evidence with different levels of reliability [35]. In this method, the assessors are asked to identify the focal sets that comprise of a single or group of criteria. The experts determine the criteria contained in the focal sets in such a way that they are able to compare them (the focal sets), given their knowledge. Then, pairwise comparison matrices are constructed for the focal sets. Using focal sets instead of single criteria allows taking into account the partial uncertainty between possible criteria. The basic belief assignments (BBA) of the corresponding focal sets are derived from the pairwise comparison matrices. The BBAs from different experts are combined using the Dempster fusion rule. The weights for each criterion are assumed to be BBA of the corresponding focal element (single criterion), and are derived based on the maximum belief-plausibility principle in Dempster-Shafer theory, or on the maximum subjective probability obtained by probabilistic transformations using the transferable belief model [35]. Again, note that in this work, this method is applied only to derive the relative weights of the criteria, rather than using it to rank alternatives. The procedure for calculating the weights of the leaf attributes based on DST-AHP is presented below.

Constructing pairwise comparison matrices

First, the experts are asked to construct pairwise comparison matrices (also known as knowledge matrices) to compare the relative importance of the attributes and sub-attributes in the same level of the hierarchy with respect to their parent attribute. For example, the pairwise comparison matrix for the attribute modeling fidelity (T_1) is a 3×3 matrix that compares the relative importance of the modeling' fidelity daughter attributes:

$$\begin{bmatrix} 1 & MF_{12} & MF_{13} \\ MF_{21} & 1 & MF_{23} \\ MF_{31} & MF_{32} & 1 \end{bmatrix} \tag{8.12}$$

where the columns correspond to the pairwise comparisons of the daughter attributes: robustness of the results ($T_{1,1}$), suitability of the selected model ($T_{1,2}$), and quality of the application ($T_{1,3}$), respectively. The element MF_{ij} is assigned by assessing the relative importance of attribute i to attribute j following the scoring protocols in (Saaty, 2008). For example, the element MF_{12} is assigned by comparing the relative importance of $T_{1,1}$ to $T_{1,2}$.

Compared to conventional AHP comparison matrices, the expert is free to choose, based on his/her belief, the elements of the pairwise comparison matrix. These elements can be focal elements that represent a single criteria, e.g., $\{A\}$ or a distinct group of criteria, e.g., $\{A, B\}$ that are comparable favorably (to the best of expert's knowledge)

to the universal set that contains all the criteria, which allows accounting for the uncertainty in the judgment [73]. For example, the expert can choose a focal set of $\{SoM, QAp\}$ if he/she believes that it can be compared favorably to the universal set$\{SoM, QAp, RoR\}$; i.e., the set of $\{SoM, QAp\}$ can be compared to $\{SoM, QAp, RoR\}$ (the sub-attributes *SoM, QAp, RoR* were defined in Table 1-4 of [18]). Then, the expert is asked to fill the pairwise comparison matrices to represent his/her belief in the relative importance of a given set (of one or multiple attributes) compared to the others. Favoring the universal set $\{SoM, QAp, RoR\}$ over $\{SoM, QAp\}$, means that the universal set contains an element that is not contained in the other set, and at the same time it is more important than the elements of the other set, i.e., RoR is more important than SoM and QAp. Finally, as in the conventional AHP method, the consistencies of the matrixes need to be tested and the assessors are asked to update their results if the consistency is lower than the required value.

Computing the weights

In this step, the weights are derived using the conventional AHP technique, according to which the normalized principal eigenvector of the matrix represents the weights. A good approximation for solving the eigenvector problem in case of high consistency is to normalize the columns of the matrix and, then, average the rows for obtaining the weights. For more details on AHP and deriving the weights from pairwise comparison matrices, the reader might refer to [122]. Please note that, as mentioned earlier, the weights derived from the pairwise comparison matrices are assumed to be the BBA of the associated focal sets.

Reliability discounting

Usually, multiple experts are involved in evaluating the weights. Each expert is regarded as an evidence source. Reliability of an evidence source represents its ability to provide correct measures of the considered problem [73]. Shafer's reliability discounting is often used to consider the reliability of the source information in DST-AHP [125]:

$$m_{\delta}(A) = \begin{cases} \delta \cdot m(A), \forall A \subseteq \Theta, A \neq \Theta, \\ (1-\delta) + \delta \cdot m(\Theta), A = \Theta, \end{cases} \tag{8.13}$$

where Θ represents the complete set of criteria, A is the focal element in the power set 2^{Θ}, $m(A)$ is the BBA for A, $m_{\delta}(A)$ is the discounted BBA, δ is the reliability factor. A value of $\delta = 1$ means that the source is fully reliable and a value of $\delta = 0$ means that the source is fully unreliable. The reliability factor of the experts is determined by the decision maker, based on their previous knowledge and experience.

Combination of experts opinions

Next, Dempster's rule of combination [125] is used to combine two independent pieces of evidence assigned by different experts. The discounted BBAs from different experts are combined by [73]:

$$m_{1,2}^{\delta}(C) = \left(m_1^{\delta} \oplus m_2^{\delta}\right)(C) = \begin{cases} 0, C = \phi, \\ \dfrac{1}{1-K} \cdot \displaystyle\sum_{A \cap B = C \neq \phi} m_1^{\delta}(A) \cdot m_2^{\delta}(B), \ C \neq \phi, \end{cases} \tag{8.14}$$

where $m_{1,2}^{\delta}(C)$ is the new BBA resulting from the combination of the two discounted BBA $m_1^{\delta}(A)$and $m_2^{\delta}(B)$ of the two experts. K is the conflict factor in the opinions of experts and given by:

$$K = \sum_{A \cap B = \phi} m_1^{\delta}(A) \cdot m_2^{\delta}(B) \tag{8.15}$$

Pignistic probability transformation

The belief functions resulted from the discounting and combination are defined for focal sets (might contain one or multiple leaf attributes). To obtain the weights of each leaf attribute, the masses ($m_{1,2}^{\delta}(C)$) assigned to the focal sets need to be transformed into masses for the basic elements. In this work, the transferable belief model proposed by [129] is used for the transformation. In this method, the masses $m_{1,2}^{\delta}(C)$ on the credal level are converted to the pignistic level using the insufficient reason principle [129]:

$$w(x) = \sum_{C \subseteq \Theta, C \neq \phi} \frac{m(C)}{1 - m(\phi)} \frac{1_{C(x)}}{|C|}, \forall x \in \Theta \tag{8.16}$$

where $w(x)$denotes the belief assignment of a single element (x) on the pignistic level, 1_C is the indicator function of C: $1_C = 1, if x \in C and 0 otherwise.$ $|A|$ is the length of A (the number of elements in the focal set). The mass functions obtained from the pignistic probability transformation represent the relative "believed weights" of the attributes.

After obtaining the local weights of the leaf attributes with respect to their parent attribute, the global weights with respect to the top-level attribute, i.e., the trustworthiness, need to be determined. This can be done by multiplying the weight of the daughter attribute by the weights of the upper parent attributes in each level. For example, the "global weight" of the historical use with respect to the trustworthiness, denoted by $W_{global}(HU)$, is calculated by:

$$W_{global}(HU) = w(HU) \times w(SoM) \times w(MF) \tag{8.17}$$

where $w(HU), w(SoM) and w(MF)$are the local weights of the historical use, the suitability of the model, and the modeling fidelity. For simplicity reasons, hereafter the global weights for the leaf attributes are denoted by W_i and

in the framework of Figure 8.9, we have $i = 1, 2, \cdots, 27$.

8.4.3 Evaluation of the risk considering trustworthiness levels

After evaluating the level of trustworthiness for the PRA of a given hazard group, the next question is how to integrate the estimated risk from the PRA with the level of trustworthiness. In this paper, we develop a Bayesian averaging model for integrating the trustworthiness based on the "weighted posterior" method [47]. Let us consider two scenarios: the risk assessment is trustable, denoted by E_T, and its complement, i.e., the risk assessment is not trustable (E_{NT}). The risk after the integration can, then, be calculated as:

$$Risk|T = P(E_T) \cdot Risk|E_T + (1 - P(E_T)) \cdot Risk|E_{NT} \tag{8.18}$$

where $Risk|T$ is the estimation of risk after considering the trustworthiness of the PRA; $P(E_T)$ is the subjective probability that E_T will occur and is dependent on the trustworthiness of the risk assessment; $Risk|E_T$ is the estimated risk from the PRA. Due to the presence of epistemic (parametric) uncertainty in the analysis, $Risk|E_T$ is often expressed as a subjective probability distribution of the risk index. $Risk|E_{NT}$ is an alternate distribution of the risk when the decision maker thinks the PRA is not trustable. In this paper, we assume $Risk|E_{NT}$ is a uniform distribution in [0,1], indicating no preference on the value of the risk index. Similar models have been used in literature to consider unexpected events in risk analysis [77].

Determining the probability of trusting the PRA

The probability $P(E_T)$ in Eq. (8.18), which represents the decision maker's belief that the risk assessment results are correct and accurate, needs to be elicited from the decision makers. The elicitation process needs to be organized and structured to ensure the quality of the elicitation.

Different methods can be found in the literature for the assessment of a single probability using experts elicitation, such as probability wheels, lotteries betting, etc. [68]. In this work, we choose the "certainty equivalent gambles" for the elicitation. Before presenting the procedure for this method, some general recommendations need to be followed to ensure the quality of the elicitation process [68]:

1. Background and preparation: uncertain events need to be defined clearly.

2. Identification and recruitment of experts: The experts who are conducting the elicitation are chosen carefully with low-value ladenness, and a preference of being both substantively and normatively skilled.

3. Motivating experts: the purpose and use of the work need to be explained to the experts, to motivate them for the elicitation.

4. Structuring and decomposition: the dependencies and functional relationships need to be first identified by the client and agreed on and modified by the experts if necessary.

5. Probability and assessment training: the experts need to be trained to elicit probabilities.

6. Probability elicitation and verification: the expert needs to elicit the probabilities paying caution to zero values, cognitive biases, etc. After making the elicitation, the expert needs to make a summary of the elicitation and verify its adequacy.

Then, a "certainty equivalent gamble" is designed to elicit the probability of trust:

1. The decision maker is asked to compare two scenarios: (1) he/she participates in a gamble (given the information from the PRA model) where he/she wins $\$1{,}000$ if an accident occurs and $\$0$ if the accident does not occur; (2) he/she wins $\$x$ for sure.

2. The experts exchange information between them and discuss.

3. Suppose that a PRA was conducted and predicted that the consequences occur for sure, and the trustworthiness of the PRA is one of the five levels defined in Section 3.1. Then, for each level of trustworthiness, the elicitor varies the value of x until the decision maker feels indifferent between the two scenarios.

4. The probability of trust at the current level of trustworthiness is, then, calculated by:

$$p = \frac{x}{1000} \tag{8.19}$$

   where 1000 here represents the $\$1000$ that the expert gains if the accident occurs (the model prediction is correct).

5. The elicitor fits a suitable function to the five data points, in order to determine the probability of trust for trustworthiness levels between the defined levels. The shape of the fitted function should be determined based on the assessors' behavior towards taking risk in trusting a low fidelity PRA:

   - A convex function should be chosen if the assessor is risk-averse, meaning that the decision maker trusts only the PRA with high levels of trustworthiness.
   - A linear function is chosen if the assessor is risk neutral.
   - A concave function is chosen if the assessor is risk-prone, meaning that although a PRA might not have a very high level of trustworthiness, the decision maker is willing to assign a high probability of trust to it.

8.4.4 MHRA considering trustworthiness levels

To aggregate risks from multiple hazard groups considering the trustworthiness of the assessment, the trustworthiness in the PRA of each single group is evaluated and integrated into the risk estimate for the corresponding hazard group first. After the integration, the risk is expressed as a subjective distribution on the probability that a given consequence will occur. Then, the estimated risk from different hazard groups is aggregated. This step can be done by simply adding the risk distributions from different hazard groups, as shown in Eq. (8.20), where $Risk_{total}$ is the total risk considering the level of trustworthiness; $(Risk_i|T)$ is the risk from the hazard group i given the level of trustworthiness; $n$is the number of hazard groups. Monte-Carlo simulations can be used to approximate the distribution of $Risk_{total}$.

$$Risk_{total} = \sum_{i=1}^{n} (Risk_i|T). \tag{8.20}$$

8.4.5 Application

The developed framework is applied on a case study for two hazard groups in the nuclear industry: the external flooding and internal events hazard groups. The PRA models of the two hazard groups were developed and provided by Electricité De France (EDF). The level of trustworthiness is assessed for each hazard group. The risk distributions from each hazard group are, then, recalculated considering the level of trustworthiness. Finally, the risk is aggregated from the two hazard groups. Details of this applications can be found in our paper [18].

The results of the MHRA are presented in Figure 8.10. The empirical probability density function of the risk is evaluated through a Monte-Carlo simulation of 10^5 samples. As a comparison, the MHRA is also conducted using the conventional methods by adding the risk indexes from the two hazard groups directly, without considering the trustworthiness, as shown in Figure 8.10 (a). The mean value of the total risk from the two hazard groups considering the level of trustworthiness is found to be 1.303×10^{-1} (reactor· year) $^{-1}$ compared to 1.622×10^{-1} (reactor· year) $^{-1}$ without considering the level of trustworthiness. As discussed earlier, the aggregation of the risks from the two hazard groups needs to consider the different levels of trustworthiness to yield a mathematically appropriate process and a physically meaningful results. In fact, considering the level of trustworthiness in the analysis means that we are accounting for the disbelief, shortcoming, and lack of knowledge in the analysis, which leads to a broader spread-out of the distributions and a larger risk interval. The increase of the interval, in which the risk can fall, represents in fact a more realistic risk analysis as it accounts for the ignorance in the model. The increase in the spread out of probability distribution of risk leads to a higher mean value of risk, as it takes into account the fact that the PRA models of the two hazard groups are based on different levels of trustworthiness.

8.5 Summary of major contributions

In this chapter, we presented our major works related to epistemic uncertainty quantification in risk and reliability assessment. Driven by the two research questions identified in Sect. 10.1, the main contributions of our works can be summarized as follows:

1. A unified framework is proposed to quantify epistemic uncertainty in risk and reliability based on a maturity model for epistemic uncertainty management developed by us. Compared to existing methods, the developed methods allow considering completeness, model structure and parametric uncertainty. As a result, a more comprehensive understanding of the impact of epistemic uncertainty can be achieved.

2. A classification-based framework is developed to evaluate the impact of epistemic uncertainty based on pre-defined criteria. Compared to traditional methods, the developed framework is able to capture complex mappings from the pre-defined criteria to the impact of epistemic uncertainty.

3. A new multi-hazard risk aggregation method is developed so that the impact of epistemic uncertainty can be considered in the aggregation. The developed framework represents a systematic way for enhancing the risk assessment and representing a mathematically more appropriate risk aggregation process. This is done by considering the different levels of realism on which the risk analyses of the aggregated hazard groups are based and integrating it in the risk analysis. From a practical point of view, the framework is developed in systematic and practical, procedural steps that facilitate the application of the framework to real life cases. In addition, it represents an illuminating point to better inform risk-based decision making, as it represents the degree of realism of the analysis.

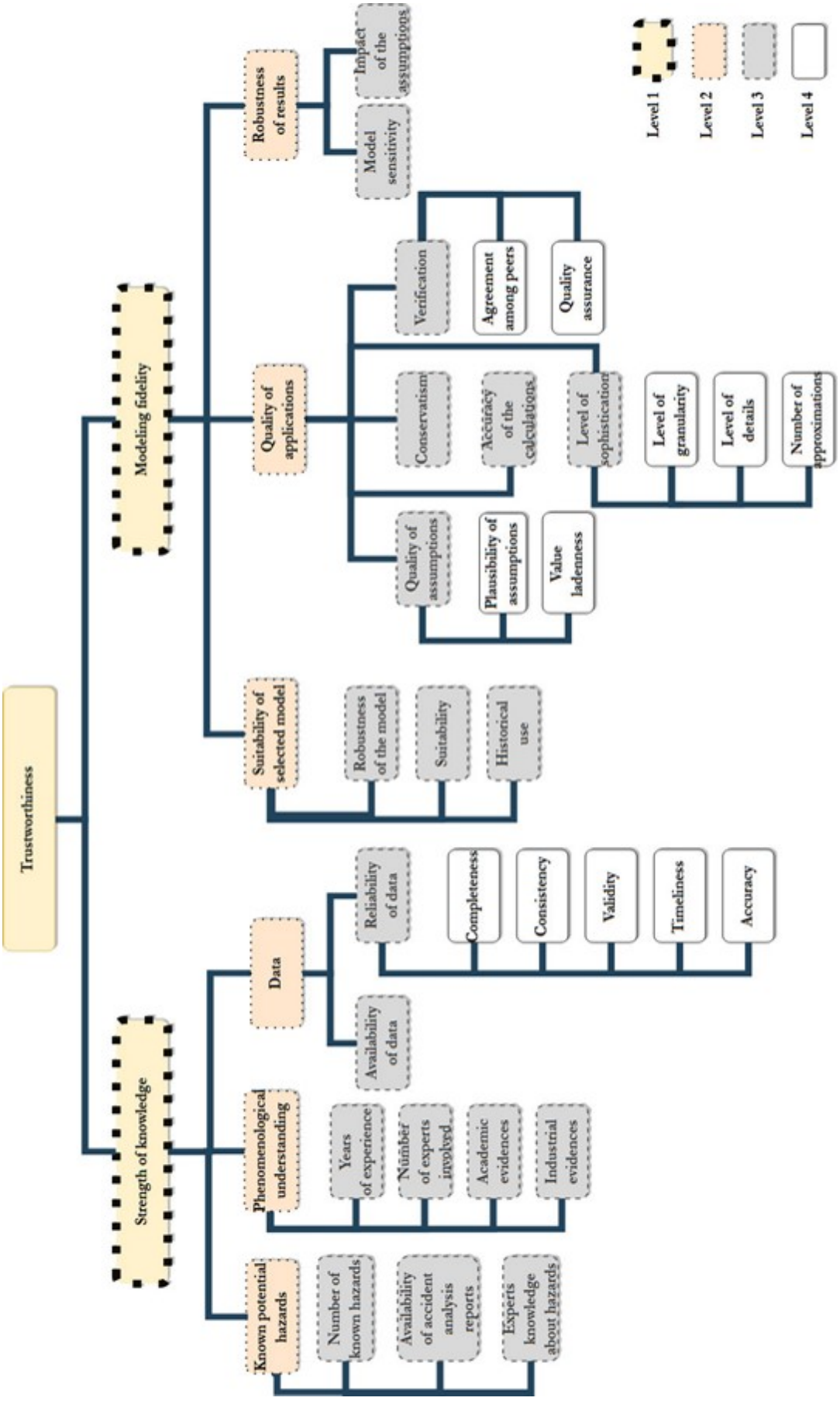


Figure 8.9: Hierarchical tree for trustworthiness evaluation.

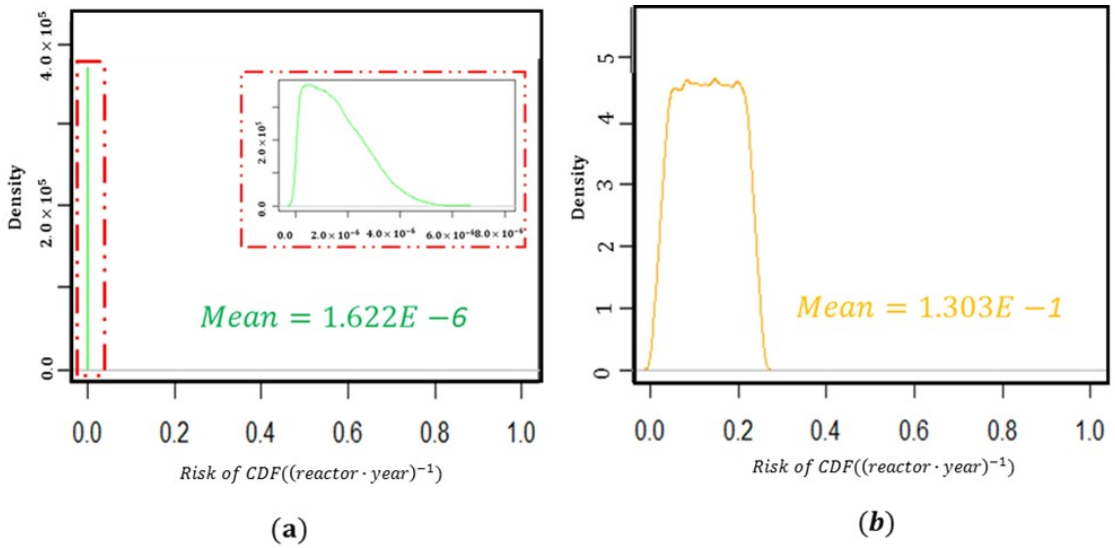


Figure 8.10: Results of the MHRA: (a) conventional aggregation, (b) considering the level of trustworthiness.

Chapter 9

MARKOV REWARD MODELS FOR RESILIENCE MODELING OF MULTI-STATE SYSTEMS

This chapter summarizes some of my representative research results in research axis 5. Unlike the previous research axes that focus only on failure behaviors, in this axis, we consider the behavior of systems that could recover from failure. Further, we consider the costs caused due to the failure and the performance losses during the recovery process. In Sect. 9.1, we briefly review the related literature and define the research problems of this research axis. A Markov reward model is developed for this and used to model the resilience of multi-state systems (Sect. 9.2). In Sect. 9.3, we extend the developed resilience model to a time-dependent case by developing a non-homogeneous semi-Markov reward process-based resilience model. Finally, in Sect. 9.4, we summarize the major contributions achieved in this research axis.

9.1 Research questions

Resilience is generally acknowledged as the ability of a system to resist, mitigate and quickly recover from potential disruptions [59]. The exact definition of resilience might differ depending on the application domain. Following a thorough review of resilience definitions from literature [162], it is concluded that a complete description of resilience should cover the following aspects:

- resistant capability, *i.e.,* the capability to resist the impact of the disruptive event and remain normal operations [58];

- absorption capability, *i.e.,* the capability to absorb the influence of the disruptive event (possibly by degrading its performance) and still remains resilient, so that the system can return to normal operation states when the disruptive event disappears [8];
- recovery capability, *i.e.,* the capability to quickly restore normal operation after the disruptive event disappears [126].

Various approaches have been developed in the literature for resilience modeling and analysis. According to the classification of Hosseini *et al.* [59], the existing approaches can be broadly classified into two categories: general measure-based and structure-based methods. In the general measure-based methods, resilience is measured based on empirically observable metrics of system performances, without considering system-specific characteristics like system structures. One of the most representative general measure-based methods is the resilience triangle developed in [24], which uses performance losses during and after disruptions to measure seismic resilience.

To use the general measure-based methods, the performance parameters need to be directly observable.This premise, however, does not always hold in practice, which limits the applicability of the general measure-based methods. Another drawback of these methods is that it does not provide explanatory models that link resilience to its contributing factors, which limits their use when design decisions need to make to improve resilience. Unlike general measure-based methods, the structure-based methods consider system-specific characteristics like system structures. Based on these system-specific characteristics, models are developed to measure resilience [59]. The structure-based methods can be further divided into optimization-based, topology-based, and simulation-based methods [135].

The optimization-based method evaluates resilience by solving an optimization model aiming at restoring the system within time constraints while minimizing the potential losses [135]. For example, Zhang *et al.* [164] developed a dynamic optimization framework to evaluate and improve the post-disaster resilience of a water distribution system. In topology-based methods, resilience is modeled and analyzed based on topological models of the systems (usually in terms of network models). This type of model is often used in vulnerability analyses, which is related to the capability of the system to resist the disruptive events and remain resilient. For example, Bose *et al.* [23] conducted a resilience and vulnerability analysis for an electrical network using the topology-based methods. Simulation-based methods use simulation to capture the uncertain behaviors involved in the resilience quantification. For example, To investigate the resilience of power systems against extreme weather events, Panteli *et al.* [108] developed a time-series Monte Carlo simulation method.

Although a substantial amount of works have been done, as reviewed above, two issues still remain to be addressed:

1. Most of the current works on resilience focus on only some of these the three aspects reviewed above. A unified and comprehensive framework for resilience quantification, which is able to consider all the aspects

mentioned above, both separately and collectively, is lacking.

2. Most of the existing methods for resilience modeling and analysis, as reviewed above, assume that the performance of the system is continuous. A lot of practical engineering systems, however, are multi-state in nature or needs to be modeled by multi-state models to control the modeling/computational complexity [19]. A typical application of multi-state models is to use them for modeling the demands and capacities of energy generation systems [72]. How to model and analyze the resilience of a multi-state system, then, becomes a problem that deserves investigation.

In this chapter, we focus on these two research questions. Section 9.2 proposes a Markov reward model for resilience of multi-state systems. The developed model is able to quantify the resistant, absorption, as well as the recovery capability of resilience. In Sect. 9.3, we extend the developed model to consider the possible time-dynamic behavior in the multi-state system. A non-homogeneous semi-Markov reward model is developed for this.

9.2 A Markov reward process-based resilience model for multistate systems

In this section, we present the developed MRP-based resilience model in Sect. 9.2.1. Then, four numerical metrics are defined in Sect. 9.2.2 for measuring resilience. In Sect. 9.2.3, we discuss how to use the developed model for resilience analysis and present a simulation-based method for evaluating the defined resilience metrics. In Sect. 9.2.4, we briefly present the application of the developed methods using a real-world case study. The work in this section is based on our publication [162]. More details could be found in the original paper.

9.2.1 A Markov reward process model for resilience

Let $X(t), t > 0$ represents the performance of a system at t under the threat of possible disruptive events. Without losing generality, let us assume that $X(t)$ takes $(m+1)$ discrete values: $X(t) \in [0, 1, 2, \cdots, m]$, where 0 represents the highest performance (perfect state) while m represents the lowest one, and that $X(t)$ is a continuous time discrete state Markov with a transition rate matrix Q (also called intensity matrix or infinitestimal generator matrix in some literature):

$$Q = \begin{bmatrix} q_{00} & q_{01} & \cdots & q_{0m} \\ q_{10} & q_{11} & \cdots & q_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ q_{m0} & q_{m1} & \cdots & q_{mm} \end{bmatrix}$$

where $q_{i,j}, 0 \leq i, j \leq m, i \neq j$ are the rates that the system departs from state i and ends in state j and $q_{i,i} = -\sum_{j \neq i} q_{i,j}, , 0 \leq i \leq m$. At $t = 0$, it is assumed that the system is in the perfect state ($X(0) = 0$). The jumps that degrade the system's performance (from state i to state j where $i < j$) are results of damages caused by disruptive events, while the jumps that improve the performance represent recovery of the system.

Typically, disruptive events can incur two types of losses on the system: the direct losses, which are generated directly by the disruptive event and do not depend on the length of the disruption; and the indirect losses, which are caused by the degraded system performances and depend on the length of the recovery process (*e.g.,* downtime costs) [151]. Take an NPP as an example. When an earthquake occurs, damages might be caused to the NPP as a direct result of the earthquake shake (*e.g.,* structural damages to the NPP, failure of components). The losses associated with these damages are called direct losses. After the earthquake, the NPP might be shut down for repairs. Financial losses are also incurred during this shutdown period due to the lost potential revenues. This kind of losses is an example of indirect losses.

To model the losses caused by the extreme events, we introduce the MRP model in Figure 9.1: the system suffers a direct loss of $d_{i,j}$ when it jumps from state i to state j due to the disruptive event, where

$$\begin{cases} d_{i,j} > 0, \text{ if } i < j, \\ d_{i,j} = 0, \text{ if } i \geq j. \end{cases} \tag{9.1}$$

Besides, the system also suffers an indirect loss of l_i (per unit of time) for its sojourn in the performance degradation state $i, 1 \leq i \leq m$.

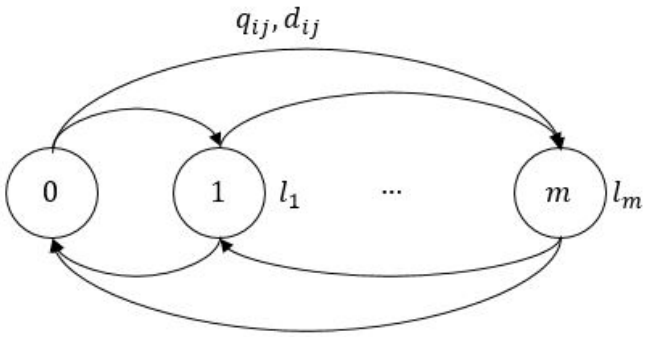


Figure 9.1: Markov reward model for resilience against extreme events.

9.2.2 Resilience metrics

As shown in Sect. 9.1, resilience of a system includes requirements on the resistant, absorption and recovery capabilities. In the following, we propose formal definitions and numerical metrics for the three aspects of resilience individually, and then propose a collective numerical metric to quantify the overall resilience of the system of interest.

Definition 6 (Resistant resilience)**.** *Resistant resilience is the ability of a system to resist the influence of extreme events without degrading its performance.*

As shown in Definition 6, resistant resilience requires the system to remain operational without performance degradations after being hit by the extreme event. In other words, a system with high resistant resilience is able to operate at full capacity after the extreme event, without the need of being repaired. Resistant resilience is often achieved through strengthening system designs, *e.g.,* strengthening structure strengths, selecting highly reliable components.

Based on the MRP model in Sect. 9.2.1, we define a numerical metric, called resistant probability (p_{Rs}), to measure the resistant resilience of a system at time t.

Definition 7 (Resistant probability)**.** *Resistant probability at time t is defined as the probability that the system can be operated at perfect performance in $(0, t)$.*

From Definition 7, p_{RS} can be calculated by

$$p_{Rs}(t) = Pr\left(X(\tau) = X_0,\ \forall \tau \in (0, t)\right), \tag{9.2}$$

where X_0 is the state with perfect system performance level. The physical meaning of p_{Rs} is the probability that the system is able to resist the impact of the extreme event. It is easy to see that p_{Rs} takes values in $[0, 1]$ and that a larger value of p_{Rs} indicates better resistant resilience. It should be noted that if we regard the event $X(t) \neq X_0$ as system failure, resilient probability is equivalent to the reliability of the system (probability of no system failure up to time t), which, according to some researchers, is an important contributor to system resilience [146].

Definition 8 (Absorption resilience)**.** *Absorption resilience is the capacity of a system to absorb the impact of extreme events so that it can be recovered to normal operation state after the extreme event vanishes, without causing permanent damages to the system.*

Absorption resilience is less demanding compared to the resistant resilience. Performance degradation is allowed as long as the impact of the extreme events can be absorbed so that the system remains in resilient states. Resilient states represent the states without permanent damages, so that the system is recoverable after the extreme events disappear. In contract, in some states, the system loses resilience. For example, an NPP attacked by an earthquake loses its resilience if the safety systems fail to promptly shutdown the NPP and a core meltdown accident occurs, like what happens in the Fukushima or Chernobyl accident. In both cases, the system loses resilience as the NPPs have to be abandoned and cannot be repaired.

To measure the absorption resilience, let us first group the state space of $X(t)$ into two classes: B_0, which contains all the resilient states, and B_1, which includes all states in which the system loses its resilience (core

meltdown accidents in NPPs, complete broken down of dams by flooding, *etc.*). Then, a numerical metric, called resilient probability (p_{Re}), can be defined to measure the absorption resilience:

Definition 9 (Resilient probability)**.** *Resilient probability at time t is defined as the conditional probability that the system remains resilient up to time t, given that disruptions occurred before t:*

$$p_{Re}(t) = Pr\left(X(t) \in B_0 \mid X(\tau) > 0, \exists \tau \in (0,t)\right). \tag{9.3}$$

It should be noted that as the non-resilient states are unrecoverable, we only need to require that $X(t) \in B_0$, rather than $X(\tau) \in B_0, \forall \tau \in (0,t)$. The physical meaning of p_{Re} is the probability that the system is able to absorb the impact of the extreme event (possibly with performance degradation) without losing resilience. It is easy to see that p_{Re} takes values in $[0,1]$ and that a larger value of p_{Re} indicates better absorption resilience. It should be noted that if we regard the states in B_1 as an undesired consequence in conventional risk analyses, p_{Re} is equivalent to the non-occurrence probability of such consequence. In engineering practice, safety barriers are often designed to prevent the system from entering the loss-of-resilience states. For example, in NPPs, a number of safety barriers (high pressure coolant injection system, automatic depressurization system, low pressure coolant injection system, *etc.*) are used in a defence-in-depth architecture to prevent severe consequences like core meltdown from happening. Adding safety barriers like these can help reduce p_{Re} and improve the absorption resilience.

Definition 10 (Recovery resilience)**.** *Recovery resilience is the capacity of a system to recover to normal operation state within required time limits after its performance is disrupted by the extreme event.*

As shown in Definition 10, recovery resilience is about whether a system can be repaired promptly within a prescribed time limit. In practice, recovery resilience depends largely on the distribution of the time needed to recover the system, which further depends on factors like maintenance resources prepared for the system, training of the maintenance personnel, *etc.* A numerical metric, called recovery probability (p_{Rc}), is defined to measure the recovery resilience:

Definition 11 (Recovery probability)**.** *Recovery probability at time t is defined as the conditional probability that the system operated in $(0,t)$ is recovered to normal operation state within a prescribed time limit $T_{th,Rc}$, given that its performance is disrupted by an extreme event.*

Let us define a random variable $T_i(t)$ to represent the accumulated sojourn time at state $i, i = 0, 1, \cdots, m$ in $(0,t)$:

$$T_i(t) = \int_0^t \mathbb{1}\left\{X(u) = i\right\} \mathrm{d}u, \tag{9.4}$$

where $\mathbb{1}\{X(u) = i\}$ is an indicator function:

$$\mathbb{1}\{X(u) = i\} = \begin{cases} 1, \text{ if } X(u) = i, \\ 0, \text{ otherwise.} \end{cases} \tag{9.5}$$

Then, p_{Rc} can be calculated by:

$$p_{Rc}(t) = Pr\left(T_{Rc}(t) \leq T_{th,Rc} \mid X(\tau) > 0, \exists \tau \in (0,t)\right). \tag{9.6}$$

where $T_{th,Rc}$ is the prescribed time threshold for system recovery; $T_{Rc}(t)$ is the accumulated recovery time in $(0,t)$ and is given by

$$T_{Rc}(t) = \sum_{i>0} T_i(t). \tag{9.7}$$

The physical meaning of p_{Rc} is the probability that the system is able to recover to normal operation states within required time limits. It is easy to see that p_{Re} takes values in $[0,1]$ and that a large value of p_{Rc} indicates better recovery resilience. Similar metrics have been seen in literature to measure the resilience from a recovery capability-based perspective. For example, in [52], resilience is measured by the conditional probability that a failed item will be recovered in the next time step, which is equivalent to Eq. (9.6) if we considered $T_{th,Rc}$ to be "the next time step".

Definition 12 (Overall resilience)**.** *Overall resilience is the capacity of a system to sustain external and internal disruptions without degrading its performance or, if the performance is degraded, to fully recover the function rapidly after the disruption vanishes.*

Overall resilience integrates the resistant, absorption and recovery resilience and provides a more complete description of system resilience. Similar definitions can also be found in literature (*e.g.,* [48], [134] and [127]). To quantitatively measure the overall resilience, let us first note that the resistant, absorption and recovery resilience can be naturally integrated through the potential losses suffered by the system:

$$\begin{aligned} L(t) &= L_D(t) + L_{ID}(t) \\ &= \sum_{i=0}^{m}\sum_{j=0}^{m} d_{i,j} \cdot N_{i,j}(t) + \sum_{i=0}^{m} l_i \cdot T_i(t), \end{aligned} \tag{9.8}$$

where $L_D(t), L_{ID}(t)$ and $L(t)$ are the direct, indirect and total loss in $(0,t)$, respectively; $N_{i,j}(t)$ is the number of system transitions from state i to j in in $(0,t)$; $d_{i,j}$ and l_i are defined in Figure 9.1 while $T_i(t)$ is defined in Eq. (9.4). In the above definition, the resistant and absorption resilience affect the direct losses, while the recovery resilience mostly affects the indirect loss. Assume that a resilience objective is set in such a way that the potential loss for the system operating in $[0,t]$ should not exceed a prescribed value of L_{tol}. Then, a numerical metric for the overall

resilience, called overall resilience metric (Re), can be defined.

Definition 13 (Overall resilience metric)**.** *Overall resilience metric at time t is defined as the probability that the potential losses caused by extreme events are within the tolerable loss L_{tol}:*

$$Re(t) = Pr(L(t) < L_{tol}). \tag{9.9}$$

The physical meaning of Re is the probability that the system does not suffer financial losses higher than a predefined threshold value L_{tol}. It is easy to see that Re takes values in $[0, 1]$ and that a larger value of Re indicates better overall resilience. The idea of using losses to quantify resilience has been adopted by various researchers. For example, it is easy to verify from Figure 9.2 that if we set $d_{i,j} = 0$ and $l_i = m - i, i, j = 0, 1, \cdots, m$, the total loss in Eq. (9.8) (the shaded area in Figure 9.2) is equivalent to the resilience triangle defined in [24]. The expected value of $L(t)$ has been widely used as a reliability metric [88], and also as a resilience metric recently [120], for electrical power system. Similar metrics are found in areas similar to resilience, *e.g.,* business continuity modelling [151], performability analysis [104]. In this paper, we also call Re overall resilience for simplicity if no confusion will be caused.

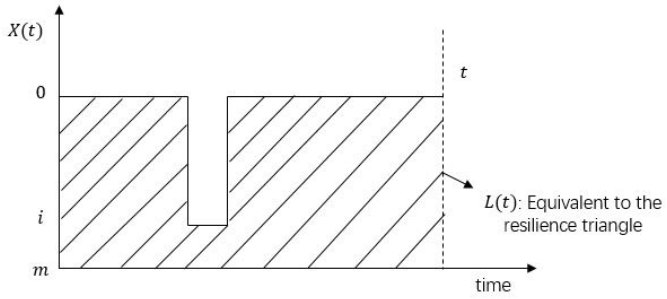


Figure 9.2: A sample trajectory of $X(t)$ and $L(t)$ with $d_{i,j} = 0$ and $l_i = m - i$.

9.2.3 Resilience modelling and analysis against the extreme events

Figure 9.3 depicts a typical event sequence after the system is hit by an extreme event. In the response phase, the built-in safety systems are activated to contain the damage caused by the extreme event. Depending on the performance of the safety systems, different consequences with different degree of damages can be resulted. After the extreme event vanishes, efforts are made to recover the system to normal operation state. Depending on the severity of consequence and also on the maintainability of the system, the required time to recovery might differ significantly.

Homogeneous Poisson processes are widely used in literature for modelling extreme events such as earthquakes [9], floods [132], hurricanes [78], etc. In this paper, we assume that the severity of the extreme event can be

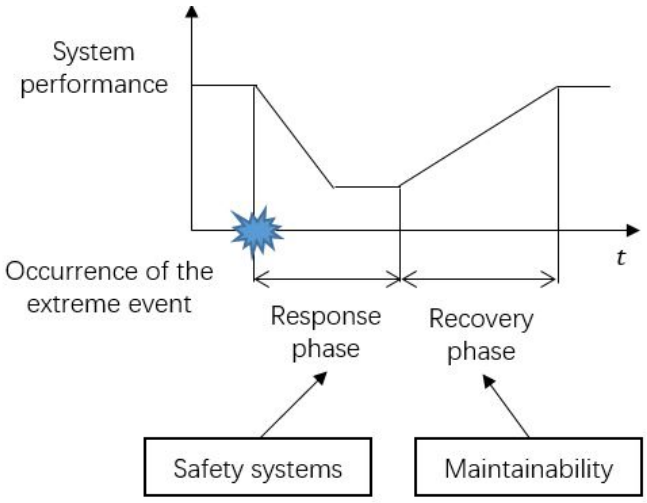


Figure 9.3: An illustration of the event sequence after the extreme event.

classified into n_S discrete levels, and the occurrence of an extreme event with severity level $S = S_i$ is modelled by a homogeneous Poisson process with a rate $\lambda_{S,i}, i = 1, 2, \cdots, n_S$. The values of $\lambda_{S,i}$ can be estimated from historical data. For example, [17] proposed a method to estimate the discretized values of $\lambda_{S,i}$ for earthquakes based on historical data and an empirical relationship called Gutenberg-Richter relationship.

Once the extreme event occurs, the system's performance might degrade, depending on the performance of the safety systems. Probabilistic combinational models, such as event trees, fault trees, binary decision diagrams, etc. [21], can be used to describe the performance of the safety systems and calculate the conditional probability for the system to be in each performance degradation state, given that an extreme event with a certain severity occurs, as shown in Figure 9.4. It is well known that the split and merge of Poisson processes are also Poisson processes [42]. Therefore, the occurrence of each system state $X = i, i = 0, 1, \cdots, m$ can be modelled by a homogeneous Poisson process with a rate λ_i, which is given by

$$\lambda_i = \sum_{j=1}^{n_S} \lambda_{S,j} \cdot Pr\left(X = i \mid S = j\right), 0 \leq i \leq m. \tag{9.10}$$

Without losing generality, we make the following assumptions:

1. states $X = 0, 1, \cdots, m-1$ are resilient states while state $X = m$ is a non-resilient state (absorbing state), *i.e.,* the system cannot be recovered if entering this state;

2. the time required to recover from state i to state j $(i > j)$ follows an exponential distribution with a rate $\mu_{i,j}$;

3. there are no damages caused by extreme events during the recovery processes.

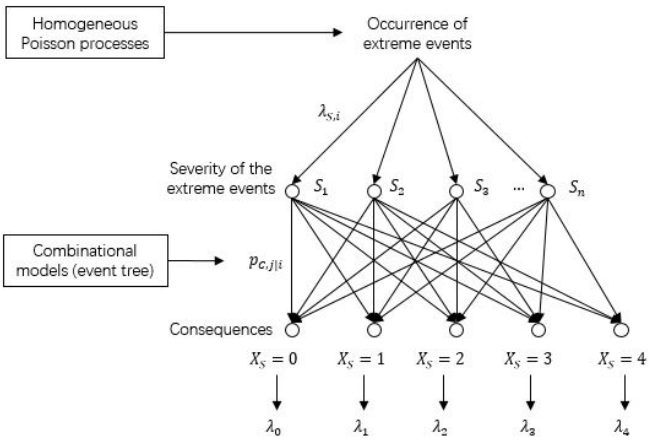


Figure 9.4: System states after the disruptive events.

Then, a MRP model defined in Sect. 9.2.1 can be established with the Q-matrix given by:

$$Q = \begin{bmatrix} -\sum_{i=1}^{m} \lambda_i & \lambda_1 & \lambda_2 & \lambda_3 & \lambda_4 & \dots & \lambda_m \\ \mu_{10} & -\mu_{10} & 0 & 0 & 0 & \dots & 0 \\ \mu_{20} & \mu_{21} & -\sum_{j=0}^{1} \mu_{2,j} & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \mu_{i0} & \mu_{i1} & \dots & -\sum_{j=0}^{i-1} \mu_{i,j} & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & 0 & \dots & 0 \end{bmatrix}. \tag{9.11}$$

The zeros in the last row indicates that the state $X = m$ is an absorbing state. The direct ($d_{i,j}$) and indirect losses (l_i) associated with the system states can, then, be determined from historical data.

A simulation method is, then, designed to calculate the resilience metrics, as shown in Algorithm 9.2. In Algorithm 9.2, N_S is the sample size of the simulation and $X = m$ indicates the state where the system loses resilience. The meaning of the other parameters can be found in the nomenclature. The algorithm used uniformization techniques [121] to generate the next state jumps. As shown in subfunction FnNextJump, the arrival time for the next jump is generated based on the largest element in each row of Q, while the next state is sampled with a probability proportional to the associated elements in Q. Once the sample paths are generated, the resilience metrics can be easily calculated by counting the direct and indirect losses. The confidence interval with a confidence level α is estimated by [42]:

$$\left[\hat{p} - Z_{1-\alpha/2} \cdot \hat{\sigma}, \hat{p} + Z_{1-\alpha/2} \cdot \hat{\sigma}\right],$$

where $\hat{p}$ is the estimated probabilities (p_{Rs}, p_{Re}, p_{Rc} and Re), Z_θ is the θ percentile of a standard normal distribution and $\hat{\sigma}$ is estimated by:

$$\hat{\sigma} = \sqrt{\frac{1}{N(N-1)} \left(n(1-\hat{p}^2) - (N_S - n)\hat{p}^2\right)},$$

Algorithm 9.1: Resilience analysis based on Markov reward model

input : $Q, d_{i,j}, l_i$
output: $p_{Rs}(t)$, $p_{Re}(t)$, $p_{Rc}(t)$, $Re(t)$

$n_{Rs} = 0$, $n_{Ab} = 0$, $n_{Rc} = 0$, $n_{Re} = 0$;
for $i \leftarrow 1$ **to** N_S **do**
  Set x_{prev}, x_{cur}, τ, τ_{next}, L_D, L_{ID}, t_{Rc} to zeros;
  while $\tau < t$ **do**
    if $x_{cur} \neq m$ **then** $L_D, L_{ID}, t_{Rc} \leftarrow$ `FnUpdateStates`;
    else break;
    $x_{prev} \leftarrow x_{cur}$;
    $x_{cur}, \tau_{next} \leftarrow$ Simulate the next jump of the Markov model using `FnNextJump`;
    $\tau \leftarrow \tau + \tau_{next}$;
  end
  if $\tau == \tau_{next}$ **then** $n_{Rs} = n_{Rs} + 1$;
  if $x_{cur} \neq m$ **then** $n_{Ab} = n_{Ab} + 1$;
  if $t_{Rc} < T_{th,Rc}$ && $t_{Rc} > 0$ **then** $n_{Rc} = n_{Rc} + 1$;
  if $L_D + L_{ID} < L_{tol}$ **then** $n_{Re} = n_{Re} + 1$;
end
$p_{Rs}(t) \leftarrow n_{Rs}/N_S$, $p_{Re}(t) \leftarrow n_{Ab}/(N_S - n_{Rs})$, $p_{Rc}(t) \leftarrow n_{Rc}/(N_S - n_{Rs})$, $Re(t) \leftarrow n_{Re}/N_S$;
Calculate the confidence intervals.
Function `FnNextJump`(x_{prev}, Q)
  output: x_{cur}, τ_{next}
  $\lambda \leftarrow -1 \cdot Q(x_{prev}, x_{prev})$;
  $\tau_{next} \leftarrow$ Generate a random number from $Exponential(\lambda)$;
  $p_i \leftarrow Q(x_{prev}, i)/\lambda$, $i = 0, 1, \cdots, m, i \neq x_{prev}$;
  $x_{cur} \leftarrow$ Generate a random number where $x_{cur} = i$ with a probability p_i;
end
Function `FnUpdateStates`$(x_{prev}, x_{cur}, \tau_{next}, d_{i,j}, l_i, L_D, L_{ID}, t_{Rc})$
  output: L_D, L_{ID}, t_{Rc}
  if $x_{prev} < x_{cur}$ **then** $L_D = L_D + d_{x_{prev}, x_{cur}}$;
  if $x_{prev} > x_{cur}$ **then** $L_{ID} = L_{ID} + l_{x_{prev}} \cdot \tau_{next}$;
  if $x_{prev} \neq 0$ **then** $t_{Rc} = t_{Rc} + \tau_{next}$;
end

where n is the number of occurrence of the associated event and N_S is the sample size.

9.2.4 Application

Resilience model

In this section, we apply the developed methods for resilience analysis of a nuclear power plant under the threat of earthquakes. A complete description of the case study can be found in Sect. IV of our paper [161]. In particular, To model the recovery process after the earthquake, we make the following assumptions:

1. the repair resource can support repairing only one NPP unit at a time;
2. if the two units both fail, unit 2 is repaired before unit 1;
3. the time required to repair one NPP unit (either 1 or 2) follows an exponential distribution with a mean value of 1.32 (years);

4. no damages are caused by earthquakes during the repair period of the NPP.

The repair sequence defined in Assumption 2 is due to the fact that unit 2 has a larger generation capacity than unit 1. The mean repair time in Assumption 3 is estimated based on data from [7]. It should be noted that the time includes both repair time and the time required for evaluation and re-licensing from the nuclear administrative. Then, the behavior of the NPP under the threat of earthquakes can be modeled by a MRP model, as shown in Figure 9.5. In Figure 9.5, the transition rates $\lambda_{0,j} = \lambda_j$; the repair rate $\mu = 1/1.32 = 0.76$ (year^{-1}). The value of the direct losses $d_{0,j}, j = 1, 2, 3$ are estimated based on the replacement cost data of NPPs in [28]. The values of the unit indirect losses $l_i, i = 0, 1, 2, 3$ are estimated based on the average electricity price data for house hold users in Europe area given in [2]. The parameter values are summarized in Table 9.1.

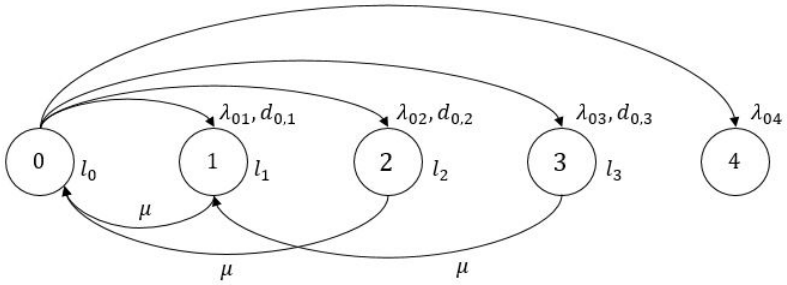


Figure 9.5: Markov reward model for the NPP.

Table 9.1: Parameter values of the Markov reward model.

Parameter	Meaning	Value	Source
$d_{0,1}$	Direct loss caused by the failure of unit 1.	1.8×10^8 (€)	Estimated using the data from [7]
$d_{0,2}$	Direct loss caused by the failure of unit 2.	1.8×10^8 (€)	Estimated using the data from [7]
$d_{0,3}$	Direct loss caused by the failure of unit 1 and 2.	3.6×10^8 (€)	Estimated using the data from [7]
$d_{0,4}$	Direct loss caused by core meltdown.	3.6×10^{10} (€)	Assumed
l_0	Indirect loss (downtime cost) per unit time for staying in state 0.	0 (€)	Estimated using the data from [28]
l_1	Indirect loss (downtime cost) per unit time for staying in state 1.	7.24×10^8 (€/year)	Estimated using the data from [28]
l_2	Indirect loss (downtime cost) per unit time for staying in state 2.	1.82×10^9 (€/year)	Estimated using the data from [28]
l_3	Indirect loss (downtime cost) per unit time for staying in state 3.	2.54×10^9 (€/year)	Estimated using the data from [28]

Results and discussions: fixed time $t = 40$ (years)

The Q-matrix of the MRP model in Figure 9.5 is

$$Q = \begin{bmatrix} -8.0629 \times 10^{-4} & 1.2298 \times 10^{-4} & 1.2298 \times 10^{-4} & 2.5939 \times 10^{-4} & 3.0095 \times 10^{-4} \\ 0.76 & -0.76 & 0 & 0 & 0 \\ 0.76 & 0 & -0.76 & 0 & 0 \\ 0 & 0.76 & 0 & -0.76 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \quad (9.12)$$

Algorithm 9.2 is used to evaluate the resilience of the NPP for a time horizon of 40 years, which is the designed life of the NPP. The sample size of the analysis is 10^6. The tolerable loss L_{tol} is assumed to be 2.54×10^9 (€) and the acceptable time limit for recovery is assumed to be $T_{Th,Rc} = 2$ (years). The point estimates of the four resilience metrics are presented in Figure 9.6 and the confidence intervals with confidence level $\alpha = 0.05$ is given in Table 9.2.

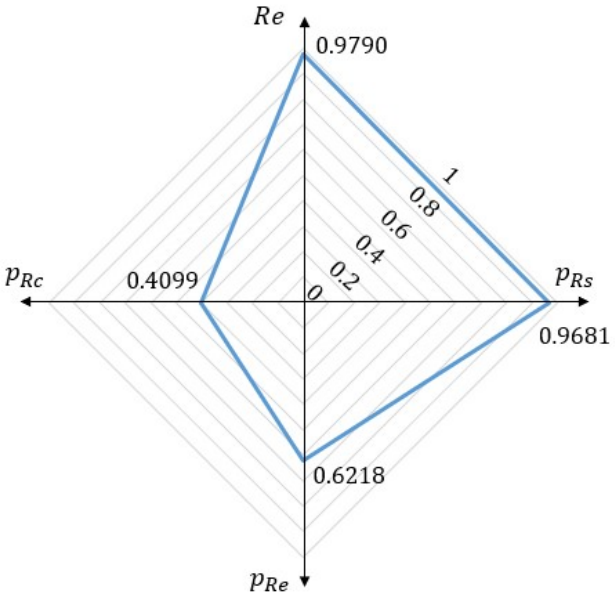


Figure 9.6: Results of the resilience analysis ($T = 40$ (years)).

Table 9.2: Confidence intervals with $\alpha = 0.05$.

	p_{Rs}	p_{Re}	p_{Rc}	Re
Lower bound	0.9678	0.6202	0.4045	0.9787
Upper bound	0.9685	0.6233	0.4153	0.9793

It can be seen from Table 9.2 that the confidence intervals are narrow. This indicates that due to the large sample size used (10^6), the estimates are accurate for supporting decision making. The results in Figure 9.6 describe different aspects of resilience for an NPP being operated up to $t = 40$ (years). For resistant resilience, we have $p_{Rs} = 0.9681,$ which indicates that one could have a high degree of belief that the generation capacity of the NPP will not be disrupted at all by earthquakes in its entire life cycle (40 years). In other words, the probability that NPP keeps operating continuously in the entire evaluation horizon without performance degradation is 0.9681. This is because the design of the NPP and its safety systems is strong for resisting the damages caused by the earthquake. As can be seen in the previous analysis, the failure probabilities of the safety systems remains at a low levels for earthquake magnitudes up to 8.5.

For the absorption resilience, we have $p_{Re} = 0.6218$. This means that if initial disruptions have already occurred, there is only a conditional probability of 0.6218 that the system remains in resilient state in the evaluation horizon, *i.e.,* no core meltdown accidents happen so that the NPP can be repaired after possible performance disruptions

caused by the earthquakes. This value might not seem satisfactory, as the core meltdown accident has a very high severity but its probability of occurrence is not low enough. To improve the absorption resilience, two possible approaches might be adopted. The first is to lower the probability of failure of the safety barriers caused by the earthquake by strengthening the anti-seismic designs. The second is to add redundant safety systems to the NPP.

For the recovery resilience, we have $p_{Rc} = 0.4099$, which indicates that there is only a probability of 0.4099 that the recovery time of the NPP can meet its requirements (recovery time should be less than $T_{Th,Rc} = 2$ (years)). This value is far from satisfactory and indicates that the recovery resilience of this NPP needs improvements. A straightforward way to improve the recovery resilience is to reduce the time-to-repair needed under each performance degradation state. There are a number of ways to achieve this, *e.g.,* providing better training to the maintenance personnel, preparing enough resources for the recovery of the NPP. Besides, the measures that improve the resistant and absorption resilience might also improve the recovery resilience. This is because the probability of entering the states with very severe performance degradations (and also requiring very long repair times) can be reduced by improving the resistant and absorption resilience.

For the overall resilience, we have $Re = 0.9790$. This means that there is a probability of 0.9790 that the total losses caused by the earthquake in the evaluation horizon do not exceed $L_{tol} = 2.54 \times 10^9$ (€). As indicated by Re, the NPP demonstrates high overall resilience as the potential losses caused by the earthquake is far below the maximal tolerable losses.

To have a complete picture of the resilience, the four resilience metrics should be considered together. As can be in Figure 9.6, although the resistant and overall resilience of the NPP are acceptable, its absorption and recovery resilience still need improvement. Hence, efforts are needed to reduce the likelihood that the NPP enters the non-resilient state (core meltdown) and to reduce the needed time to recover the NPP from performance degradation states.

9.3 A non-homogeneous Semi-Markov reward process-based resilience model

In this section, we present the developed non-homogeneous Semi-Markov reward model to consider the time-dependent behavior in the multi-state systems. The model is developed in Sect. 9.3.1. Procedures for applying the developed model is, then, introduced in Sect. 9.3.2. In Sect. 9.3.3, a numerical algorithm is developed for efficient resilience assessment based on the developed model. The performance of the model and the evaluation algorithm are tested in Sect. 9.3.4. The work in this section is based on our publication [161]. More details could be found in the original paper.

9.3.1 The model

Let S be a discrete (finite or countable) space (called state space hereafter) in which a discrete random variable X_n takes values. Let θ_n be a continuous random variable taking values in $[0, \infty)$ and let $\tau_n = \sum_{i=0}^{n} \theta_n$, $n = 1, 2, \cdots$ Physically, θ_n is often used to model the inter-arrival time between the nth and $(n-1)$th state transition in X_n, and τ_n represents the accumulated time up to the nth state transition. A bivariate process $\{X_n, \tau_n\}$, $n = 0, 1, 2, \cdots$ is called a non-homogeneous renewal process (NHRP) when the following assumptions hold [46]:

$$\begin{aligned} &Pr(X_{n+1} = j, \tau_{n+1} - \tau_n \leq t \mid X_n = i, \tau_n = \tau, X_{n-1}, \tau_{n-1}, \cdots, X_0, \tau_0) = \\ &Pr(X_{n+1} = j, \tau_{n+1} - \tau_n \leq t \mid X_n = i, \tau_n = \tau). \end{aligned} \tag{9.13}$$

and

$$Pr(X_0 = i, \tau_0 = 0) = Pr(X_0 = i). \tag{9.14}$$

Let $N(t) = \sup_n\{\tau_n \leq t\}$, $n = 0, 1, \cdots$ Then, a stochastic process $\{X(t), t \geq 0\}$ with piecewise constant and right continuous sample paths given by $X(t) = X_{N(t)}$ is called a non-homogeneous semi-Markov process (NHSMP) associated with the NHRP $\{X_n, \tau_n\}$ [46]. It is called NHSMP as it is a semi-Markov process with reward structure. A semi-Markov process is called "semi", as it has Markov property only at instants when state transitions occur. In an NHSMP model, future system behaviors depend not only on the current state but also on the age of the system, which makes it an ideal tool to capture time-dependent system behaviors.

Let us assume that the performance of the system can be characterized by an NHSMP $X(t)$ with $n+1$ performance levels: $X(t) \in S = \{0, 1, 2, \cdots, n\}$, as shown in Figure 9.7. Without loss of generality, we assume that the system performance decrease as the value of $X(t)$ increases (*i.e.,* $X(t) = 0$ indicates perfect performance while $X(t) = n$ indicates the worst performance). Disruptive events might impair system performances and cause the system to jump from a higher performance level to a lower performance level. After the disruptions, recovery measures can be taken to restore the system performances, resulting in backward jumps from lower performance levels to higher ones. The NHSMP can be determined by initial distribution and renewal kernel [46]. Initial distribution is a probability vector representing the distribution of states at $t = 0$. Let $\vec{\pi}_0 = [Pr(X(0) = i) : i \in S]$ represent the initial probability vector of the NHSMP. In this paper, we assume that at $t = 0$, the system starts operation from the perfect performance state. Hence, we have $\vec{\pi}_0 = [1, 0, \cdots, 0]$. Let $Q(t, \tau) = [Q_{i,j}(t, \tau) : i, j \in S]$ represent the renewal kernel, where

$$Q_{i,j}(t, \tau) = Pr(X_{n+1} = j, \theta_{n+1} \leq t \mid X_n = i, \tau_n = \tau). \tag{9.15}$$

In Eq. (9.15), τ_n represents the time when the nth state transition occurs; θ_{n+1} is the inter arrival time between the $(n+1)$th and the nth state transition: $\theta_{n+1} = \tau_{n+1} - \tau_n$, where τ_{n+1} is the time when the $(n+1)$th state transition occurs. The physical meaning of the renewal kernel $Q_{i,j}(t, \tau)$ is the joint probability that the next state is j and the

time to next transition is no greater than t, given that the current state is i and that the age up to the nth transition is τ.

When rewards can be accumulated when system jumps between states and/or stays in a state, a NHSMP becomes a NHSMRP. In this paper, we use a NHSMRP to characterize the resilience of the system of interest, as shown in Figure 9.7, where the rewards represent the direct and indirect losses caused by disruptive events. Direct losses are generated directly by disruptive events and do not depend on the length of the disruption. The degraded system performances cause indirect losses before the system is fully recovered (*e.g.,* revenue losses) [151]. Hence, indirect loss depends on the length of the recovery process. In this paper, reward rates $d_{i,j}$ represents the direct loss suffered by the system when it degrades from state i to state j. It is easy to verify that

$$\begin{cases} d_{i,j} > 0, \text{ if } i < j, \\ d_{i,j} = 0, \text{ if } i \geq j, \end{cases} \tag{9.16}$$

as $i < j$ indicates performance degradation, while $i \geq j$ indicates recovery. Similarly, the indirect losses per unit of time of sojourn in the performance degraded state $i, i = 1, 2, \cdots, n$ are modeled by reward rate l_i. Please note that $l_0 = 0$ as state 0 represent perfect performance and there is no indirect loss for this state.

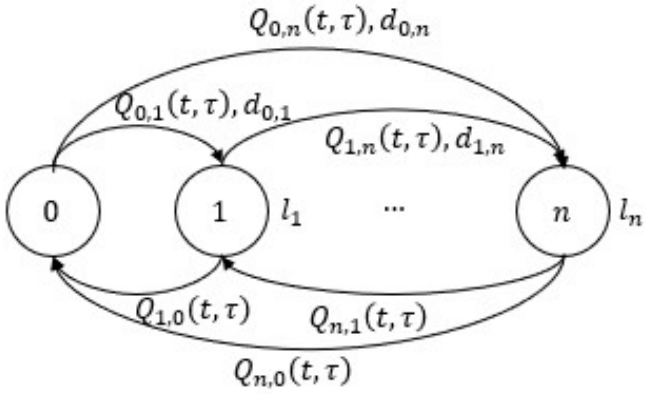


Figure 9.7: An illustration of the NHSMRP-based resilience model.

The four resilience metrics defined in Sect. 9.2.1 can be naturally extended to quantify resilience based on the NHSMRP model in Figure 9.7.

9.3.2 Procedures of applying the model

Figure 9.8 summarizes the procedures for applying the developed model for resilience modeling and analysis. The analysis starts with identifying the disruptive events that might threaten the system of interest. Typical disruptive events considered in the resilience analysis of energy systems include extreme weather, failure of components, natural disasters, *etc.* In this paper, we limit our analysis to the case where the system is subject to the threat of

only one disruptive event. The developed methods can also be extended naturally to systems subject to multiple disruptive events. Homogeneous Poisson processes are widely used for modeling the occurrence of disruptive events like earthquakes, floods, hurricanes, *etc* [78]. As these existing works, we also assume that the occurrence of the disruptive event follows a homogeneous Poisson process with a rate λ_D. The value of λ_D can be estimated from historical data or based on expert judgment [17].

Various safety barriers protect the system from the disruptive event. Depending on the performance of these safety barriers, different consequences can result following the occurrence of a disruptive event. For example, when an earthquake occurs, safety barriers like NPP structure strengthening, emergency trip system, emergency cooling systems, *etc.,* are activated to protect a NPP from severe consequence like core meltdown. Depending on whether these safety barriers successfully perform their designed function, different consequences can result, *e.g.,* (in ascending order of severity), unaffected operation, operation interruption, core meltdown. In the second step of the analysis, a consequence analysis determines the possible consequences that might be caused by the disruptive event and the losses associated with each consequence. The losses include both direct and indirect losses: the former are determined directly by estimating the financial losses caused to the system by the disruptive event, while the latter is determined by estimating the cost per unit of time of sojourn in the performance degraded states. Each consequence is mapped into a state in the NHSMRP model (see Figure 9.7). The values of $d_{i,j}$ and l_i, $i, j = 0, 1, \cdots, n$ are determined based on the estimated losses caused by the corresponding consequences.

Then, a probabilistic analysis calculates the occurrence probability of each consequence. When the performance of the safety barriers does not change with time, the analysis can be easily done by combining the Poisson process model with probabilistic combinational models like event tree [39]. In this paper, we consider safety barriers with time-dependent performances, for which the existing models cannot be directly applied. In Sect. **??**, we develop an analytical approach to calculate the probability density function of the occurrence time for each consequence.

The next step is estimating the recovery time. The estimation can be done by collecting field recovery time data or recovery exercise data and estimate the CDF of the time needed to recover to a given performance level. These CDFs, together with the distribution of the occurrence time for each consequence, are used to derive the renewal kernel matrix of the NHSMRP.

Following the previous steps, a NHSMRP is constructed for resilience analysis. Compactly, let us denote the developed model by a tuple $< \vec{\pi}_0, Q(t,\tau), D, \vec{l} >$, where $D = [d_{i,j} : i, j \in S]$ is a matrix whose elements represent the corresponding direct losses and $\vec{l} = [l_0, l_1, \cdots, l_n]$ is a vector that contains the unit indirect loss for each state. Based on the developed model, the four resilience metrics defined in Sect. 9.2.2 are calculated for resilience analysis. In Sect. 9.3.3, we present an efficient Monte Carlo simulation algorithm to calculate the resilience metrics.

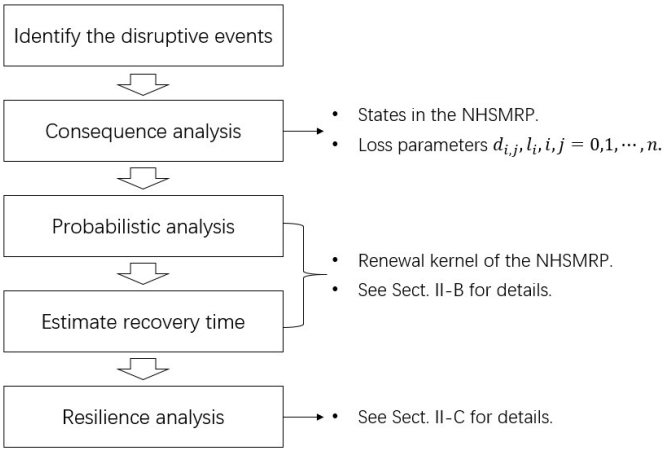


Figure 9.8: Procedures of applying the developed model for resilience analysis

9.3.3 Efficient Monte Carlo simulation for resilience analysis

Resilience analysis of the developed NHSMRP model requires analyzing its time-dependent behavior. Although there are a few existing approaches for transient analysis of an NHSMRP model, a common challenge is the high computational burden, especially when the state space of the NHSMRP gets large. In this section, we develop an efficient Monte Carlo simulation algorithm to improve the computational performance of resilience analysis. The performance of this algorithm will be discussed through both theoretical analyses and numerical experiments in the next section.

The algorithm is based on the embedded chain and holding time distributions of the NHSMP. Let $\tau_i, i = 1, 2, \cdots$ be the time instant when the ith system state transition (X_i) occurs. For an NHSMP described in Figure 9.7, it is well-known that X_i follows a non-homogeneous discrete time discrete state Markov chain with a initial distribution $\vec{\pi}_0$ and a transition probability matrix $P(\tau) = [p_{i,j}(\tau) : i, j \in S]$, where $p_{i,j}(\tau)$ is given by [46]:

$$p_{i,j}(\tau) = \lim_{t \to \infty} Q_{i,j}(t, \tau). \tag{9.17}$$

The discrete time Markov chain $X_i, i = 0, 1, \cdots$ is called the embedded chain of the NHSMP. Let $F_{i,j}(t, \tau)$ denote the CDF of the holding time $T_{i,j}$, which is defined by:

$$\begin{aligned} F_{i,j}(t, \tau) &= Pr(T_{i,j} \leq t) \\ &= Pr(\theta_{n+1} \leq t \mid X_n = i, X_{n+1} = j, \tau_n = \tau), i, j \in S, t \geq 0. \end{aligned} \tag{9.18}$$

In Eq. (9.18), θ_{n+1} denote the interarrival time between the n and $(n+1)$th state transition, and τ_n is the accumulated

time up to the nth state transition. It is easy to see that [46]

$$\begin{aligned} F_{i,j}(t,\tau) &= \frac{Pr(\theta_{n+1} \leq t, X_n = i, X_{n+1} = j, \tau_n = \tau)}{Pr(X_n = i, X_{n+1} = j, \tau_n = \tau)} \\ &= \frac{Pr(\theta_{n+1} \leq t, X_n = i \mid X_{n+1} = j, \tau_n = \tau)}{Pr(X_n = i \mid X_{n+1} = j, \tau_n = \tau)} \\ &= \frac{Q_{i,j}(t,\tau)}{p_{i,j}(\tau)}. \end{aligned} \tag{9.19}$$

The PDF of $T_{i,j}$ can also be derived:

$$\begin{aligned} f_{i,j}(t,\tau) &= \frac{dF_{i,j}(t,\tau)}{dt} \\ &= \frac{1}{p_{i,j}(\tau)} \frac{dQ_{i,j}(t,\tau)}{dt} \\ &= \frac{1}{p_{i,j}(\tau)} f_{\eta_{i,j}}(x,\tau) \prod_{k \in S, k \neq j} \left(1 - F_{\eta_{i,k} \mid \eta_{i,j}}(x,\tau)\right). \end{aligned} \tag{9.20}$$

Having derived $p_{i,j}(\tau)$ and $F_{i,j}(t,\tau)$, The NHSMRP can be easily simulated for resilience analysis. In practice, however, calculating $p_{i,j}(\tau)$ through Eq. (9.17) often requires numerical integration, as the integration in Eq. (**??**) is often too complicated to have analytical solutions. The numerical integration could bring very high computational burdens, as it has to be evaluated for each generated sample. In the developed algorithm, we attempt to solve this problem by introducing a linear interpolation model (Algorithm 9.2 - Algorithm 9.4). The developed algorithms comprise of two phases. In the training phase, a linear interpolation model is trained based on n_{tr} training samples and used to approximate the transition probability matrix $P(\tau)$ of the embedded chain $X_i, i = 0, 1, \cdots$ The trained linear interpolation model is, then, used in the simulation phase to generate state jumps. In this way, numerical integration of Eq. (9.17) can be avoided in the simulation phase. When n_{tr} is large, the linear interpolation model can approximate $P(\tau)$ fairly accurately. At the same time, the computational costs of running the linear interpolation model are much less than directly doing a numerical integration. Since $P(\tau)$ needs to be evaluated for each generated state jump, the linear interpolation model can greatly improve the computational efficiency of the algorithm. In general, a larger value of n_{tr} would have better approximation accuracy, but paying the price of increasing computational costs. In practice, a trade-off needs to be made to balance the accuracy and computational costs. Another major difference between the developed algorithms and the traditional Monte Carlo simulation is that, in the developed algorithms, samples are generated using the idea of vectorization. Instead of generating state jumps interactively using loops, vectorized functions are designed that take vector inputs and generate vectors of the state jumps directly. Vectorization can greatly reduce the overhead cost of the program, and, therefore, dramatically improve the computational efficiency of the algorithm. Algorithm 9.2 - Algorithm 9.4 also apply to Markov reward models. In this case, only $p_{i,j}(\tau)$ and $F_{i,j}(t,\tau)$ need to be replaced by their counterparts in the Markov reward models, while the algorithms can remain unchanged.

Algorithm 9.2: Resilience analysis based on the NHSMRP model.

Training phase:
$\vec{\tau}_{tr} \leftarrow$ Equally insert n_{tr} points into $[0, T]$;
for $i, j \in S$ **do**
  $\vec{p}_{tr} \leftarrow$ Evaluate Eq. (9.17) with $\tau = \vec{\tau}_{tr}$, using numerical integration methods;
  $\tilde{p}_{i,j}(\tau) \leftarrow$ Train a linear interpolation model based on training data $(\vec{\tau}_{tr}, \vec{p}_{tr})$;
end
Simulation phase:
Set $\vec{x}_{cur}, \vec{x}_{next}, \vec{\theta}, \vec{t}_{cur}, \vec{t}_{next}$ to column vectors with N_S elements of zeros;
while *At least one element in* $\vec{t}_{cur} < T$ **do**
  for $x_{cur} =$ *Unique states in* $\vec{x}_{cur}$ **do**
    $\vec{\tau} \leftarrow \vec{t}_{cur}(\vec{x}_{cur} == x_{cur})$;
    $\vec{x}_{next}(\vec{x}_{cur} == x_{cur}) \leftarrow$ Starting from state x_{cur}, generate the next state with parameter $\vec{\tau}$ (Algorithm 9.3);
    $\vec{\theta}(\vec{x}_{cur} == x_{cur}) \leftarrow$ Starting from state x_{cur} and ending with states $\vec{x}_{next}(\vec{x}_{cur} == x_{cur})$, generate holding times with parameter $\vec{\tau}$ based on Algorithm 9.4;
    $\vec{t}_{next}(\vec{x}_{cur} == x_{cur}) \leftarrow \vec{\tau} + \vec{\theta}(\vec{x}_{cur} == x_{cur})$;
    $\vec{t}_{next}(\vec{t}_{next} > T) \leftarrow T$; $\vec{x}_{next}(\vec{t}_{next} > T) \leftarrow \vec{x}_{cur}(\vec{t}_{next} > T)$;
    $\vec{L}(\vec{x}_{cur} == x_{cur}) \leftarrow$ Update the total losses based on Eq. (9.8);
  end
  $\vec{x}_{cur} \leftarrow \vec{x}_{next}$; $\vec{t}_{cur} \leftarrow \vec{t}_{next}$;
end
Estimate resilience metrics based on Eqs. (9.2) - (9.9).

Algorithm 9.3: Vectorized simulation to generate the next jump

input : $x_{cur}, \vec{\tau}, \tilde{p}_{x_{cur},j}, j = 1, 2, \cdots, n_{state}$
output: $\vec{x}_{next}$
$P = [p_{i,j}] \leftarrow \tilde{p}_{x_{cur},j}(\vec{\tau}(i)), i = 1, 2, \cdots, length(\vec{\tau}), j = 0, 1, \cdots, n$;
$C \leftarrow$ Calculate the column-wised cumulative sum of the matrix P;
$\vec{u} \leftarrow$ Generate a column vector of random numbers with the same size as $\vec{\tau}$ from $U[0, 1]$;
$\vec{x}_{next} \leftarrow$ For each row in C, find the index of the first element larger than the element in the same row of $\vec{u}$;

In Algorithm 9.2, T represent the evaluation horizon; $\vec{x}_{cur}$ and $\vec{x}_{next}$ represent the current and the next state, respectively; $\vec{t}_{cur}$ and $\vec{t}_{next}$ present the time instant for the current and the next jump, respectively; $\vec{\theta}$ represents the holding times. The logical operations on vectors, *e.g.,* $\vec{\theta} > 0$, returns a logical index vector whose element is 1 if the the logical operation on the corresponding element of $\vec{\theta}$ is true (and 0 otherwise). In programming languages like Matlab, the logical index vector can be directly used to access the corresponding elements in the original vector. For example, $\vec{\theta}(\vec{\theta} > 0)$ returns the positive elements in $\vec{\theta}$.

Algorithm 9.3 is vectorized generations of the inverse transform sampling method for generating random numbers [121]. The inputs x_{cur} is the current state; $\vec{\tau}$ is a column vector that contains the ages of the samples; $\tilde{p}_{x_{cur},j}, j = 1, 2, \cdots, n_{state}$ are the linear interpolation models to approximate the transition probabilities $p_{x_{cur},j}, j = 1, 2, \cdots, n_{state}$, respectively. The output $\vec{x}_{next}$ is a vector of the same size as $\vec{\tau}$, which contains the generated next states for samples with current state x_{cur} and age $\vec{\tau}$.

Algorithm 9.4 is a vectorized generation of the acceptance rejection sampling method [121]. The inputs $f_{i,j}(x, \tau)$

Algorithm 9.4: Generate holding times using vectorized acceptance-rejection method.

input : $f_{i,j}(x,\tau), g_{i,j}(x,\tau), c_{i,j}(\tau), \vec{\tau}$
output: $\vec{\theta}$
while $\vec{\tau}$ *is not empty* **do**
  $\vec{u} \leftarrow$ Generate a column vector of random numbers with the same size as $\vec{\tau}$ from $U[0,1]$;
  $\vec{\theta} \leftarrow$ Generate a column vector of random numbers with the same size as $\vec{\tau}$ from $g_{i,j}(x,\vec{\tau})$;
  Keep the elements in $\vec{\theta}$ which satisfy $\vec{u}. * c_{i,j}(\vec{\tau}) \leq f_{ij}(\vec{\theta},\vec{\tau})./g_{ij}(\vec{\theta},\vec{\tau})$ and discard the other elements;
  Delete the elements in $\vec{\tau}$ whose holding time has been sampled in $\vec{\theta}$;
end

is the PDF of the holding time, as defined in Eq. (9.20); $g_{i,j}(x,\tau)$ is a proposal density function which is easier to simulate; $c(\tau)$ is a function that satisfies $f_{i,j}(x,\tau) \leq c(\tau) \cdot g_{i,j}(x,\tau), \forall x$ and τ. The operators $.*$ and $./$ represent multiplication and division operators for vectors, *i.e.,* apply the corresponding operations on each element of the vectors.

9.3.4 Performance analysis and numerical experiments

Theoretical analysis

The computational complexity of the developed method, denoted by O_d, depends on the complexity of the simulation algorithm O_{sim}, and the overhead cost O_{oc} :

$$\begin{aligned} O_d &= O_{sim} + O_{oc} \\ &= O(n_s \bar{n}_t n_{st}) O_p + O(n_s \bar{n}_t) O_{ht} + O(\bar{n}_t) O_{oh}, \end{aligned} \tag{9.21}$$

where n_s is the sample size of the simulation, $\bar{n_t}$ is the average number of state jumps in $(0,t)$, n_{st} is the number of states, O_p is the computational complexity of evaluating one element of Eq. (9.17), O_{ht} is the computational complexity of generating one sample from the hitting time distribution, and O_{oh} represents the overhead cost of per loop.

In the literature, there are two types of methods for analyzing the behaviors of Markov/semi-Markov reward models: Monte Carlo simulations [96] and numerical integrations [66, 130]. Table 9.3 lists some of the most widely used methods in the literature and discusses their computational complexities. It can be seen from the Table that, compared to the numerical integration methods, a significant benefit of the developed method is that its computational complexity only grows linearly with the state numbers, while the numerical integration methods have at least quadratic growth rates. Hence, the computational performance of the developed method would be significantly better than the numerical integration methods for problems with large state space. Compared to the standard Monte Carlo simulation, the computational performance of the developed method outperforms in the two aspects: first, thanks to the linear interpolation model, the O_p of the developed methods is much lower; second, the overhead cost

is also much lower in the developed method.

Table 9.3: Similar methods in the literature.

Method	Category	Applicable models	Computational complexity
Tijms' method [130]	Numerical integration	Markov reward model	$O\left(\frac{n_{st}^2 ty}{\Delta^2}\right)$ [29]
Janssen's method [66]	Numerical integration	Semi-Markov reward model	$O\left(n_{st}^2\left(\frac{t^2}{\Delta^2}+\frac{t}{\Delta}\right)\right)$
Standard Monte Carlo [96]	Monte Carlo simulation	Both Markov and semi-Markov reward model	$O(n_s\bar{n}_t n_{st})O_p + O(n_s\bar{n}_t)O_{ht} + O(n_s)O_{oh}$

Numerical experiment - I

Two numerical studies are conducted to compares the computational performances. The first is based on a homogeneous Markov reward process model (a special case of the NHSMP) described in Sect. 4.2 of [29]. The data used in this experiment are from Table 4.2 of [29]. Let Y_t denote the accumulated reward at t. As in [29], we calculate $P(Y_{10} \leq 10)$ under different sizes of state spaces (n_{st}). Three methods are tested: the developed method, Tijms' method [130] and standard Monte Carlo simulation [96]. The parameter values, *i.e.,* the step size Δ in the Tijms' method and the sample size n_s in the other two methods, are set to be $\Delta = 0.01, n_s = 10^6$. The reason for setting parameters like this is that it can ensure the three methods achieve the same degree of errors.

The computational experiment has been carried out on a computer with a CPU of 2.59 GHz (12 cores) and 64 GB RAM. Table 9.4 presents the results of the numerical experiment. In Table 9.4, M is a parameter that defines the problem and controls the size of the state space (see [29] for details), and n_{st} is the number of states of the derived model. Theoretical values for $M = 4, 5, 6, 7$ are provided in [29] through high-precision numerical integrations and are used in this study as reference values. As in [29], we keep increasing the state space until the running time of the algorithm exceeds 300 (seconds) with $\Delta = 0.01, n_s = 10^6$. Figure 9.9 shows the comparisons of the running times in log-scales. It can be seen from the comparisons that when the state space is small, the Tijms' method performs better than the developed methods. While as the state space increases, the developed method outperforms. This is because, as discussed in the theoretical analysis, the computation time of the numerical integration methods increases quadratically with the state space, while the developed method has a linear increasing rate. The developed method performs always better than the standard Monte Carlo simulations. This is due to the benefits from the reduced overhead costs through vectorization, and also from the reduced cost of simulating next jumps through the use of linear interpolation.

Numerical experiment - II

A second numerical experiment is conducted to compare the performances of developed method with Janssen's method [66] and standard Monte Carlo simulation [96] on an NHSMRP model. We design a simple case study

Table 9.4: $Pr(Y_{10} \leq 10)$ calculated by different methods.

M	n_{st}	Developed method	Tijms' method [130]	Standard Monte Carlo [96]	Theoretical value from [29]
4	5	0.445951	0.444693	0.444687	0.434068
5	13	0.226824	0.226727	0.226679	0.214623
6	25	0.171730	0.170841	0.171006	0.160671
7	41	0.142809	0.142358	0.142764	0.137375
9	81	0.121781	0.121266	0.121818	–
30	1513	0.085572	–	0.085364	–
60	$6s613$	0.080194	–	–	–

–: Running time exceeds 300 (seconds).

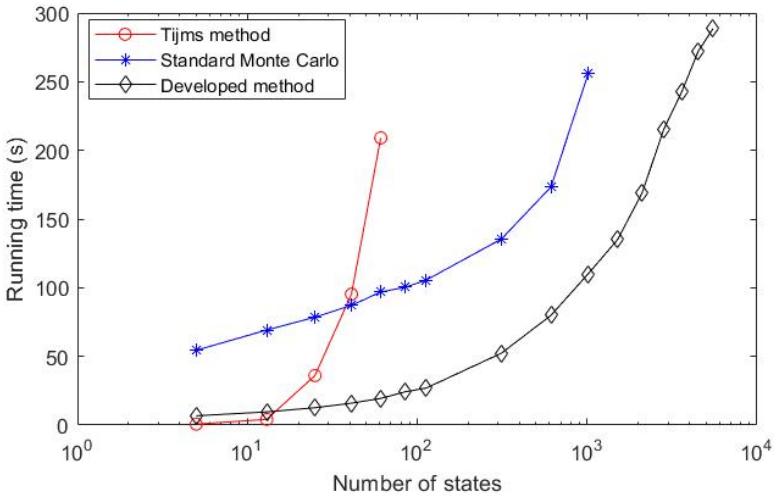


Figure 9.9: Running times of the first numerical experiment.

following the same protocol as [66]. A three-state system is defined to describe the event tree model in Figure **??** and used in this case study. The arrival of the initiating event is assumed to be a Poisson process with rate λ and p_{II} is assumed to be a time-dependent function whose values are given by the inverse CDF of a Weibull distribution with parameters η_f and β_f. The recovery time distribution from state 1 to state 0 is assumed to follow a Weibull distribution with parameters η_r and β_r. The parameter values used in the analysis are summarized in Table 9.5.

Table 9.5: Parameter values of the second numerical experiment (in arbitrary units).

Parameters	λ	p_I	η_f	β_f	η_r	β_r
Values	10^{-3}	1.6×10^{-3}	100	2	1	2

The three methods are used to calculate the state probabilities at $t = 1000 : Pr(X(1000) = i), i = 0, 1, 2$. The computational experiment has been carried out on a computer with a CPU of 2.59 GHz (12 cores) and 64 GB RAM. The results and computation times are compared in Table 9.6. For the two Monte Carlo simulation methods, the two-sided confidence intervals with a confidence level $\alpha = 0.05$ are also given. It can be seen from the comparison that even for a relatively small-scale NHSMP, the developed efficient Monte Carlo simulation performs much better

than the numerical integration method. As shown in the previous section, when the sample size grows larger, the computational benefits of the developed method over the numerical integration methods would become more significant. This is because, since the system is nonhomogeneous, considering a given time t requires considering all the history up to t, which makes numerical integration method computationally expensive. Compared to the standard Monte Carlo simulation, the developed method can achieve significant performance improvement. This confirms our theoretical analysis that the developed method improves the standard Monte Carlo simulation by reducing the computational costs in evaluating the transition probability matrix of the embedded chain through linear interpolation, and also by reducing the overhead cost through vectorization.

Table 9.6: Results of the second numerical experiment.

Methods	Developed method	Standard Monte Carlo	Janssen's method
$Pr(X(1000) = 0)$	0.996768 ± 0.000883	0.996701 ± 0.000915	0.995583
$Pr(X(1000) = 1)$	0.002349 ± 0.000111	0.002384 ± 0.000112	0.002022
$Pr(X(1000) = 2)$	0.000058 ± 0.000095	0.000059 ± 0.000096	0.002394
Computation time (s)	6.80	328.78	172.12

9.4 Summary of major contributions

In this chapter, we present our works related to resilience modeling of multi-state systems. Focusing on the two research questions identified in Sect. 9.1, the main contributions of our works can be summarized as follows:

1. A resilience model is developed for multi-state energy systems based on Markov reward process models. Four numerical metrics are defined to measure the three aspects of system resilience separately (resistant, absorption and recovery resilience) and collectively (overall resilience). The developed methods are applied for resilience analysis of a NPP against seismic hazards. The result of the analysis shows that the developed methods can provide a comprehensive description of the resilience for a given evaluation horizon.

2. A non-homogeneous Semi-Markov reward process-based model is developed for quantifying resilience of multi-state systems. The time-dependent behavior of system resilience can be captured in the developed model thanks to its non-homogeneous nature. A simulation algorithm was developed for resilience analysis based on vectorization and linear interpolation. Computational experiments showed that the developed outperforms the existing methods, thanks to the reduced overhead costs through vectorization and the reduced computational costs of generating samples through linear interpolation.

Chapter 10

MULTI-SOURCE DATA INTEGRATION FOR RELIABILITY ASSESSMENT

This chapter summarizes some of my representative research results in research axis 5. The focus of this axis is to develop a methods that support fusing data from different sources for online reliability assessment and remaining useful life prediction. In Sect. 10.1, we briefly review the related literature and define the research problems of this research axis. Some representative results are briefly introduced in Sects. 10.2 - 10.4, which dicuss the fusion of statistical failure data with condition-monitoring data, inspection data with condition-monitoring data, expert judgment with condition-monitoring data, respectively. Finally, in Sect. 10.5, we summarize the major contributions achieved in this research axis.

10.1 Research questions

Traditionally, risk and reliability assessment is conducted based on historical failure/accident data [98]. In practice, however, failure data are often scarce (if available at all), which defies the use of classical statistical methods and challenges Bayesian methods with respect to the assumption of subjective prior distributions [14]. On the other hand, in the life cycle of products, there are also other data sources that contain information regarding product risk/reliability, e.g., condition-monitoring data, inspection data, expert judgment. If used properly, these additional data sources can complement statistical failure data and improve the accuracy of risk/reliability assessment.

Condition-monitoring data are one of the most widely-used additional data source for risk/reliability assessment. Condition-monitoring data refer to the online-monitoring data related to the system's operational state and degradation processes [82]. In practice, accident initiating events and safety barriers failures usually occur as a result of degradation mechanisms, *e.g.*, wear [156], corrosion [157], fatigue [71], crack growth [20], oxidation [32], etc. These

degradation processes can be monitored and failures can be predicted and anticipated with reference to specific thresholds of the monitored variables. Condition-monitoring data contain information on the individual degradation process of the target system and provide the opportunity to update the reliability values before actual failures occur. There are a few initial attempts of using condition-monitoring data in DRA. For example, Zadakbar *et al.* applied Kalman filtering to estimate the true degradation states from condition-monitoring data and conducted DRA based on a loss function associated with the degradation states [148]. Similar works were also conducted by the same authors using different condition-monitoring techniques, *i.e.*, Particle Filtering (PF) [149] and Principal Component Analysis (PCA) [147]. The works reviewed above use only condition-monitoring data for risk updating, and do not consider statistical failure data. How to integrate condition-monitoring data with statistical failure data, then, remains a challenge for a more informed DRA.

Inspection data are collected by physical inspections performed by maintenance personnel [105] and have also been widely used as an additional data source for online reliability assessment. For example, a Bayesian method has been developed to merge experts' judgment with continuous and discontinuous inspection data for the reliability assessment of multi-state systems [93]. A two-stage recursive Bayesian approach has been developed in [92], in order to update system reliability based on imperfect inspection data. Condition monitoring data and inspection data on wind turbine blades have been used separately for remaining useful life estimation in [106]. As inspections directly measure the component degradation, they provide valuable information complementary to condition monitoring data for DRA and can help reducing the impact of the uncertainty in the condition monitoring data on the result of DRA. However, to the best of our knowledge, no previous work has considered integrating condition monitoring data and inspection data for DRA.

Apart from condition-monitoring and inspection data, some subjective knowledge on the health state of a system can also be collected from expert judgment. For example, the health state of bearings can be evaluated by experts via direct visual inspections or indirect measurements [141]. In aviation, experts might be able to evaluate the health state of turbofan engines during the breaks between two adjacent missions [118]. Expert knowledge can also provide insight into the health state and could be integrated with the CM information to achieve more accurate prognostics. However, as pointed out by Si et al. [128] and Lei et al. [84], the effective use of subjective expert knowledge for RUL prediction remains an open challenge. More specifically, the challenges include: (1) to quantify expert knowledge imprecision due to the vagueness of expert judgments and/or the measurement uncertainty; (2) to fuse two different types of information, i.e., expert knowledge and CM information. Existing literature has made some attempts on these challenges. For example, He et al. [53] introduced an exponential model to characterize the degradation of Li-ion batteries, where the model parameters were initialized by combining different imprecise expert knowledge. However, they did not use expert knowledge in the operation phase of Li-ion batteries to support RUL prediction. Ramasso and Denoeux [118] developed a partially-hidden Markov model (PHMM) to estimate model parameters by combining expert knowledge and observations. They found that including expert knowledge drasti-

cally improved the performance of parameter estimation. Nevertheless, the PHMM assumed that observations are discrete, and only used one state sequence in the offline training phase. Such a model cannot be straightforwardly implemented on continuous CM information from non-repairable systems. To the best of our knowledge, existing models did not fuse expert knowledge and CM information for RUL prediction.

As shown in the reviews above, various data sources could be used to complement statistical failure data for a better risk/reliability assessment. However, most of the existing researches consider the different data sources separately. How to fuse the information from multiple heterogeneous data sources, is, then, a challenging issue for online reliability assessment and remaining useful life prediction.

10.2 Fusing statistical failure data and condition-monitoring degradation data for dynamic risk assessment

Statistical failure data refer to count data of failures, accidents or near misses from similar systems [100, 99]. Condition-monitoring data come from online monitoring the degradation of the target system of interest [82]. In this section, we develop a Bayesian updating algorithm that integrates Particle Filtering (PF) with Markov Chain Monte Carlo (MCMC) to fuse the statistical and condition-monitoring data for online reliability and risk assessment. The work in this section was previous published as a journal paper [152]. More information could be found there.

The problem we intend to address in this paper is formally defined in Subsection 10.2.1. A first step in the DRA is to online update the reliability of the safety barriers using the two types of data. For this, a hierarchical Bayesian reliability model is developed in Subsection 10.2.2. Based on the model, an online assessment algorithm is developed for the reliability values of the safety barriers in Subsection 10.2.3 and 10.2.4. A sequential Bayesian algorithm is developed in Subsection 10.2.5 to update the risk indexes using the revised reliability values of the safety barriers. Finally, in Subsection 10.2.6, we show an application of the developed method on a High-Flow Safety System (HFSS) and compare its result with a benchmark model from literature.

10.2.1 Problem definition

Without loss of generality, we consider an ET with n possible consequences $C_1, C_2, \cdots, C_n$, m safety barriers $B_1, B_2, \cdots, B_m$ and an initial event IE. Conceptually, the ET can be expressed as

$$\mathbf{r}_C = g_{\mathsf{ETA}}(R_{B_1}, R_{B_2}, \cdots, R_{B_m} \mid \mathsf{IE}), \tag{10.1}$$

where R_{B_i} is the reliability of the ith safety barrier and $\mathbf{r}_C = [r_{C_1}, r_{C_2}, \cdots, r_{C_n}]$ is the consequence risk index considered in this paper, which is measured by the conditional occurrence probability of the consequence given that

IE has occurred:

$$r_{C_i} = Pr\{C_i \mid \text{IE has occurred}\}, i = 1, 2, \cdots, n. \tag{10.2}$$

In this paper, we consider the dynamic assessment of the risk indexes as defined in (10.1), using both statistical failure data and condition-monitoring data. Statistical data refer to the count data of the consequences of accidents that occur during the operation of similar systems, thus providing "population" information, while condition-monitoring data come from online monitoring the degradation of the specific target system of interest and describe system-specific features. More specifically, it is assumed that:

1. statistical failure data and condition-monitoring data are collected at predefined observation instants $t = t_j, j = 1, 2, \cdots, q$;

2. the collected statistical failure data are denoted by $N_{k,j}, k = 1, 2, \cdots, n$, where $N_{k,j}$ denotes the number of the kth consequences that occur in the interval $(t_{j-1}, t_j]$ and $t_0 = 0$;

3. the collected condition-monitoring data on the ith safety barrier at $t = t_j$ are denoted by $y_{i,j}$, $i = 1, 2, \cdots, m$ and $j = 1, 2, \cdots, q$;

4. the degradation threshold for the ith safety barrier is $y_{th,i}$ and failure of the ith safety barrier occurs when $y_{i,j} \leq y_{th,i}$.

The problem of DRA can, then, be defined as: at each $t = t_j, j = 1, 2, \cdots, q$, update the estimation of $\mathbf{r}_C$ in (10.1), based on statistical failure data $N_{k,j}$ and condition-monitoring data $y_{i,j}$.

10.2.2 Hierarchical Bayesian model for safety barrier reliability updating

In this section, a hierarchical Bayesian model is developed for evaluating the reliability of the safety barriers considering both statistical and condition-monitoring data. The model is based on the following assumptions:

1. in each interval $(t_{j-1}, t_j], j = 1, 2, \cdots, q$, the ith safety barrier in the population of similar systems has reliability $\pi_{i,j}$, where $\pi_{i,j}$ is a random variable with prior distribution $p_{0,\pi_{i,j}}$ and posterior distribution $p_{1,\pi_{i,j}}$ and $t_0 = 0$;

2. the prior distribution of $\pi_{i,1}$ is a Beta distribution with parameter α_i and β_i :

$$\pi_{i,1} \sim \text{Beta}(\alpha_i, \beta_i) \tag{10.3}$$

while for $j \geq 2$, $p_{0,\pi_{i,j}} = p_{1,\pi_{i,j-1}}$;

3. in each interval $(t_{j-1}, t_j], j = 1, 2, \cdots, q$, the reliability of the ith safety barrier in the target system of interest, denoted by $R_{B,i,j}$, is a random variable whose prior distribution is a Beta distribution:

$$R_{B,i,j} \sim \text{Beta}\left(K\pi_{i,j}, K(1-\pi_{i,j})\right), \tag{10.4}$$

where $\pi_{i,j}$ follows its posterior distribution $p_{1,\pi_{i,j}}$.

4. K is a random variable with uniform prior distribution:

$$K \sim \text{Uniform}(K_L, K_U). \tag{10.5}$$

From Assumption 1, the statistical count data of occurrence of accidents with given consequences in each interval can be modeled by a binomial probability model:

$$Pr\left\{N_{S,i,j}, N_{F,i,j} \mid \pi_{i,j}\right\} \propto \pi_{i,j}^{N_{S,i,j}}(1-\pi_{i,j})^{N_{F,i,j}}, \tag{10.6}$$

where $N_{S,i,j}$ and $N_{F,i,j}$ represent the number of successes and failures of the ith safety barrier in $(t_{j-1}, t_j]$, respectively. The detailed procedures for calculating $N_{S,i,j}$ and $N_{F,i,j}$ from the statistical failure data are given in Subsection 10.2.3. The reason for us to choose the binomial model is that the statistical failure data on the safety barriers are of failure-on-demand type [81, 50]. Equation (10.6) serves as the likelihood function for the statistical failure data. It should be noted that, for simplicity, we drop the constants in the likelihood function, since they do not affect the derivation of posterior distributions in Bayes theorem [80].

According to Assumption 2, at each $t_j, j = 1, 2, \cdots, q$, the prior distribution in (10.3) can be updated recursively based on Bayesian theorem [50]. Since the likelihood function in (10.6) is conjugate to the Beta prior in (10.3), the posterior $p_{1,\pi_{i,j}}$ is also a Beta distribution [50]:

$$\pi_{i,j} \sim \text{Beta}\left(\alpha_i + \sum_{\tau=1}^{j} N_{S,i,\tau}, \beta_i + \sum_{\tau=1}^{j} N_{F,i,\tau}\right). \tag{10.7}$$

Assumption 3 relates the condition-monitoring data to the statistical failure data. To explain it, note that the mean value of the Beta distribution in (10.4) is calculated by [50]:

$$E\left[R_{B,i,j}\right] = \frac{K\pi_{i,j}}{K\pi_{i,j} + K(1-\pi_{i,j})} = \pi_{i,j}.$$

Therefore, it is assumed that statistical failure data from similar systems determine the mean value of the reliability of the target system under condition-monitoring. Let $M_{S,i,j}$ and $M_{F,i,j}$ denote the number of successes and failures of the ith safety barrier in $(t_{j-1}, t_j]$, respectively. Assumption 3 also indicates that $M_{S,i,j}$ and $M_{F,i,j}$ can be modeled

by a binomial model:

$$Pr\{M_{S,i,j}, M_{F,i,j} \mid R_{B,i,j}\} \propto R_{B,i,j}^{M_{S,i,j}}(1-R_{B,i,j})^{M_{F,i,j}}. \tag{10.8}$$

Note that $M_{S,i,j}$ and $M_{F,i,j}$ have to be generated from condition-monitoring data by conducting "pseudo-tests", since in practice, we have only one sample, *i.e.*, that of the target system under condition-monitoring. Detailed procedures of generating $M_{S,i,j}$ and $M_{F,i,j}$ are discussed in details in Subsection 10.2.3. Equation (10.8) is the likelihood function for the condition-monitoring data. As in (10.6), the constants in the likelihood function are dropped since they do not affect the derivation of the posterior distributions [50].

As discussed in [50], K can be regarded as the "prior sample size". Let $M = M_{S,i,j} + M_{F,i,j}$ denote the sample size of the pseudo-tests based on the condition-monitoring data. Roughly speaking, the ratio between K and M measures the trust on the statistical failure data compared to the condition-monitoring data: a high value of K/M indicates that one has more trust on the statistical failure data than the condition-monitoring data, and vice versa. In practice, the value of K should be determined based on the value of M to reflect the weight of trust on the two types of data. A detailed discussion on the effect of K is given in Section **??**. Assumption 4 accounts for the uncertainty in determining the precise value of K.

Some existing works can be found in literature for DRA, e.g., [100] *etc.* These models, however, do not assume a hierarchical structure for the reliability and, therefore, can only be used for modeling statistical failure data. Compared to the existing models, the uniqueness of the developed model is that it proposes a hierarchical Bayesian model, which allows integrating both statistical failure data and condition-monitoring data.

10.2.3 Generating pseudo-test data

Pseudo-test data are an important concept in the developed DRA method. They are generated, based on the collected data, (either statistical failure data or condition-monitoring data), to represent the "equivalent" binomial tests and failure data on each safety barrier. In this paper, we distinguish two types of pseudo-tests:

Statistical data-based pseudo-tests

Statistical data ($N_{k,j}, k = 1, 2, \cdots, n, j = 1, 2, \cdots, q$) count the number of occurrences of the consequences in each observation interval. Note that in an ET, observing a certain consequence indicates that the events associated to it have occurred. Since the events correspond to success or failure of the safety barriers, the statistical failure data can be viewed as pseudo-tests on the safety barriers. Take a simple ET in Figure 10.1 as an example. From Figure 10.1, we can see that if consequence C_2 occurs, safety barrier B_1 must be working and B_2 must be failed. Therefore, the occurrence of C_2 is equivalent to a pseudo-test on B_1 whose result is success and a pseudo-test on B_2 whose result is failure. The same reasoning applies to the other consequences and safety barriers. Let us define

an indicator function $\mathbb{1}(B_i, C_k)$:

$$\mathbb{1}(B_i, C_k) = \begin{cases} 1, \text{ if the occurrence of } C_k \text{ indicates the} \\ \quad \text{success of the } i\text{th safety barrier,} \\ 0, \text{ if the occurrence of } C_k \text{ indicates the} \\ \quad \text{failure of the } i\text{th safety barrier.} \end{cases} \tag{10.9}$$

The pseudo-test data $N_{S,i,j}$ and $N_{F,i,j}$ can, then, be calculated from $N_{k,j}$:

$$\begin{aligned} N_{S,i,j} &= \sum_{k=1}^{n} \mathbb{1}(B_i, C_k) \cdot N_{k,j}, \\ N_{F,i,j} &= \sum_{k=1}^{n} (1 - \mathbb{1}(B_i, C_k)) \cdot N_{k,j}. \end{aligned} \tag{10.10}$$

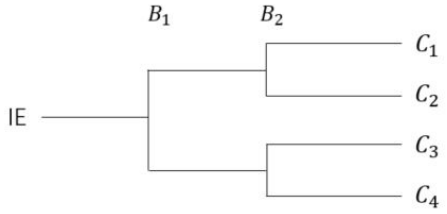


Figure 10.1: An illustrative ET

Condition-monitoring data-based pseudo-tests

Condition-monitoring data ($y_{i,j}, j = 1, 2, \cdots, q$) are collected by online-monitoring the degradation process of the ith safety barrier at $t = t_j, j = 1, 2, \cdots, q$. Since condition-monitoring data are often subject to process and observation noises, PF is used in this paper to estimate the true degradation states. PF is chosen for its flexibility and ability to handle complex nonlinear system dynamics and non-Gaussian noises. Although other methods, such as extended Kalman filter and unscented Kalman filter, might also be applied on nonlinear and non-Gaussian problems, they are based on Taylor approximation of a non-linear function. PF, on the other hand, does not require such approximation and fully represent the nonlinear system dynamics.

It is assumed that the degradation process of the ith safety barrier follows a state space model [10]:

$$\begin{cases} \mathbf{x}_{i,j} = g_i\left(\mathbf{x}_{i,j-1}, \epsilon_i\right) \text{ (state equation)}, \\ y_{i,j} = h_i\left(\mathbf{x}_{i,j}, \delta_i\right) \text{ (observation equation)}, \end{cases} \tag{10.11}$$

where $\mathbf{x}_{i,j}$ is the state variable, $y_{i,j}$ is the observation, ϵ_i is the process noise and δ_i is the observation noise.

$$p\left(\mathbf{x}_{i,j} \mid y_{i,1}, y_{i,2}, \cdots, y_{i,j}\right) = \frac{p\left(y_{i,j} \mid \mathbf{x}_{i,j}\right) p\left(\mathbf{x}_{i,j} \mid y_{i,1}, y_{i,2}, \cdots, y_{i,j-1}\right)}{\int p\left(y_{i,j} \mid \mathbf{x}_{i,j}\right) p\left(\mathbf{x}_{i,j} \mid y_{i,1}, y_{i,2}, \cdots, y_{i,j-1}\right) d\mathbf{x}_{i,j}}, \tag{10.13}$$

In PF, the forms of $g_i(\cdot)$ and $h_i(\cdot)$ are assumed to be known and the true system state $\mathbf{x}_{i,j}, j = 1, 2, \cdots, q$ are estimated recursively based on Bayesian theorem [10] (Eq. (10.13)), where in (10.13), $p\left(\mathbf{x}_{i,j} \mid y_{i,1}, y_{i,2}, \cdots, y_{i,j}\right)$ is the posterior density for $\mathbf{x}_{i,j}$, updated at $t = t_j$; $p\left(y_{i,j} \mid \mathbf{x}_{i,j}\right)$ is determined by the observation equation in (10.11) and $p\left(\mathbf{x}_{i,j} \mid y_{i,1}, y_{i,2}, \cdots, y_{i,j-1}\right)$ is determined based on the output of the PF at $t = t_{j-1}$.

In practice, (10.13) is evaluated using sequential Monte Carlo simulations: at each t_j, $p\left(\mathbf{x}_{i,j} \mid y_{i,1}, y_{i,2}, \cdots, y_{i,j}\right)$ is approximated by

$$p\left(\mathbf{x}_{i,j} \mid y_{i,1}, y_{i,2}, \cdots, y_{i,j}\right) \approx \sum_{k=1}^{N_P} w_{i,j}^{(k)} \delta\left(\mathbf{x}_{i,j} - \mathbf{x}_{i,j}^{(k)}\right) \tag{10.12}$$

where $\left\{\mathbf{x}_{i,j}^{(k)}, w_{i,j}^{(k)}\right\}, k = 1, 2, \cdots, N_P$ are the samples (referred to as "particles") and the associated weights generated by sequential importance sampling, and $\delta(\cdot)$ is the Dirac delta function.

It is shown in [10] that if at each $t = t_j$, the particles are generated by

$$\mathbf{x}_{i,j}^{(k)} \sim p(\mathbf{x}_{i,j} \mid \mathbf{x}_{i,j-1}), \tag{10.14}$$

where $p(\mathbf{x}_{i,j} \mid \mathbf{x}_{i,j-1})$ is the proposal density of the importance sampling and is determined by the state equation in (10.11), then, the weights can be updated by

$$w_{i,j}^{(k)} = \frac{w_{i,j-1}^{(k)} p\left(y_{i,j} \mid \mathbf{x}_{i,j}^{(k)}\right)}{\sum_{k=1}^{N_P} w_{i,j-1}^{(k)} p\left(y_{i,j} \mid \mathbf{x}_{i,j}^{(k)}\right)}. \tag{10.15}$$

Algorithm 10.1 [10] summarizes the major steps of the PF here employed. The purpose of resampling in Algorithm 10.1 is to avoid the well known problem of particle degeneracy and resampling is often conducted by sampling with replacement from $\left\{\mathbf{x}_{i,j-1}^{(k)}, w_{i,j-1}^{(k)}\right\}_{k=1}^{N_P}$ [60].

At each $t = t_j$, the posterior density of $\mathbf{x}_{i,j}$ is approximated by the updated particles and weights from sequential importance sampling. Therefore, the particles can be viewed as pseudo-tests on the reliability of the safety barriers, based on which $M_{S,i,j}$ and $M_{F,i,j}$ can be generated (Algorithm 10.2).

10.2.4 Updating the reliability of the safety barriers

In this section, we discuss how to update the reliability of the safety barriers based on the pseudo-test data generated in Subsection 10.2.3. The updating is done in two stages. In the first stage, statistical failure data are used to update the reliability of similar systems ($\pi_{i,j}$). As shown in Assumption 2, the prior distribution of $\pi_{i,j}$ and the statistical failure data follow a beta-binomial model [50]. Therefore, the posterior density of $\pi_{i,j}$ can be recursively updated

Algorithm 10.1: PF-based estimation of the states of the safety barriers [10]

input : $\left\{\mathbf{x}_{i,j-1}^{(k)}, w_{i,j-1}^{(k)}\right\}_{k=1}^{N_P}, y_{i,j}$

output: $\left\{\mathbf{x}_{i,j}^{(k)}, w_{i,j}^{(k)}\right\}_{k=1}^{N_P}$

for $k = 1 : N_P$ **do**

  Sample $\mathbf{x}_{i,j}^{(k)}$ using (10.14)

;

Update $w_{i,j}^{(k)}, k = 1, 2, \cdots, N_P$, using (10.15);

$\hat{N_{eff}} \leftarrow \left(\sum_{k=1}^{N_P}\left(w_{i,j}^{(k)}\right)^2\right)^{-1}$;

if $\hat{N_{eff}} < N_P/2$ **then** Update $\mathbf{x}_{i,j}^{(k)}$ and $w_{i,j}^{(k)}$ by resampling;

return $\left\{\mathbf{x}_{i,j}^{(k)}, w_{i,j}^{(k)}\right\}_{k=1}^{N_P}$.

Algorithm 10.2: Generating pseudo-test data based on PF

input : $\left\{\mathbf{x}_{i,j}^{(k)}, w_{i,j}^{(k)}\right\}_{k=1}^{N_P}, y_{i,th}$

output: $M_{S,i,j}$, $M_{F,i,j}$

$M_{S,i,j} = 0$, $M_{F,i,j} = 0$

for $k = 1 : N_P$ **do**

  $\mathbf{x}_{pseudo}^{(k)} \leftarrow$ Randomly select one element from $\left\{\mathbf{x}_{i,j}^{(k)}\right\}_{k=1}^{N_P}$, where $\mathbf{x}_{i,j}^{(k)}$ is selected with probability $w_{i,j}^{(k)}$;

  Calculate $y_{pseudo}^{(k)}$ using the observation equation in (10.11);

  if $y_{pseudo}^{(k)} > y_{i,th}$ **then** $M_{S,i,j} = M_{S,i,j} + 1$;

  else $M_{F,i,j} = M_{F,i,j} + 1$;

return $M_{S,i,j}, M_{F,i,j}$.

using (10.7). The updated posterior density is, then, combined with condition-monitoring data in the second stage to update the reliability of the safety barriers ($R_{B,i,j}$).

To do this, first note that $R_{B,i,j}$ is modeled by a hierarchical Bayesian model with a hyper-parameter K (see Assumptions (3) and (4) in Subsection 10.2.2). It should be mentioned that the $\pi_{i,j}$ in (10.4) is not regarded as a hyper-parameter, but as a random variable with a fixed probability distribution (*i.e.*, $p_{1,\pi_{i,j}}$ yielded by the first stage updating). Based on Bayes theorem [50], the joint posterior density of $R_{B,i,j}$ and K, denoted by $p_1(R_{B,i,j}, K)$, can be expressed as

$$\begin{aligned} p_1(R_{B,i,j}, K) &\triangleq p\left(R_{B,i,j}, K \mid M_{S,i,j}, M_{F,i,j}\right) \\ &\propto p\left(M_{S,i,j}, M_{F,i,j} \mid R_{B,i,j}\right) \cdot \\ &\quad p\left(R_{B,i,j} \mid K\right) \cdot p(K), \end{aligned} \tag{10.16}$$

where $p\left(M_{S,i,j}, M_{F,i,j} \mid R_{B,i,j}\right)$ is the likelihood function in (10.8), $p\left(R_{B,i,j} \mid K\right)$ is the prior distribution of $R_{B,i,j}$ in (10.4), and $p(K)$ is the prior distribution of K in (10.5). Equation (10.16) can be further expressed as (10.20) where $B(\cdot)$ is the Beta function and Δ is a proportional constant.

Due to the complexity of (10.20), it is hard to derive the analytical form of $p_1(R_{B,i,j}, K)$. Therefore, we use

Markov Chain Monte Carlo (MCMC) to generate samples from $p_1(R_{B,i,j}, K)$. For a detailed discussion of MCMC, readers might refer to Chapter 3 in [50]. Note that in this case, if we fix the value of K in (10.20), we have (10.21), which indicates that conditioned on K and the data, $R_{B,i,j}$ follows a Beta distribution:

$$\begin{aligned} R_{B,i,j} \mid\ & K, M_{S,i,j}, M_{F,i,j} \\ & \sim \text{Beta}(M_{S,i,j} + K\pi, M_{F,i,j} + K(1-\pi)). \end{aligned} \tag{10.17}$$

Therefore, in the MCMC, $R_{B,i,j}$ can be updated using Gibbs sampler based on (10.17) [50]. On the other hand, if we condition on $R_{B,i,j}$ and the data, we have:

$$\begin{aligned} & p(K \mid R_{B,i,j}, M_{S,i,j}, M_{F,i,j}) \propto \\ & \quad R_{B,i,j}^{K\pi} \cdot (1 - R_{B,i,j})^{K(1-\pi)} \cdot \frac{1}{B(K\pi, K(1-\pi))}, \end{aligned} \tag{10.18}$$

which cannot be expressed as any known probability distribution. Therefore, the Metropolis-Hastings (MH) algorithm is used to update K. In this case, we choose the proposal distribution to be a Uniform distribution over $[K_L, K_U]$, *i.e.*, the same as the prior distribution of K. Therefore, the acceptance probability p_{acc} becomes [50]:

$$\begin{aligned} p_{acc} &= \min\left(1, \frac{p(\theta^* \mid data)}{p(\theta^{(l-1)} \mid data)} \frac{f(\theta^{(l-1)} \mid \theta^*)}{f(\theta^* \mid \theta^{(l-1)})}\right) \\ &= \min\left(1, \frac{p\left(K^{(*)} \mid R_{B,i,j}, M_{S,i,j}, M_{F,i,j}\right)}{p\left(K^{(l-1)} \mid R_{B,i,j}, M_{S,i,j}, M_{F,i,j}\right)}\right), \end{aligned} \tag{10.19}$$

where $f(\cdot \mid \cdot)$ is the proposal density and the ratio in (10.19) is calculated based on (10.18).

A hybrid Gibbs/MH algorithm is developed to dynamically update the reliability of the safety barriers, as shown in Algorithm 10.3, where N_l is the number of the iterations. As l becomes large, $\left\{R^{(l)}, K^{(l)}\right\}$ converge to a random sample from the joint posterior distribution [50]. In practice, the first $N_{burn-in}$ samples are dropped to reduce the correlation between the samples [80]. Therefore, at each $t = t_j$, Algorithm 10.3 is used to update the reliability of the ith safety barrier and the posterior density of $R_{B,i,j}$ is approximated by $R^{(l)}, l = N_{burn-in} + 1, N_{burn-in} + 2, \cdots, N_l$.

One thing that needs special attention when applying Algorithm 10.3 is to check the convergence of the MCMC samples. Normally, the MCMC algorithms start from initial values that might be far away from the center of the posterior distribution. As the algorithm iterates, the MCMC samples tend to converge to samples from the posterior distribution. In this paper, we use trace plots for the convergence checks: a stable trace plot indicates good convergence, while a trace plot with significant increasing or decreasing trends means that more iterations are needed for convergence [50]. Some numerical indicators, e.g., autocorrelation coefficient, sample standard deviation of the batch means, potential scale reduction, *etc.*, can also be used to monitor the convergence of the MCMC. For more details, readers might refer to Chapter 3 of [50].

Algorithm 10.3: A hybrid Gibbs/MH algorithm to update the reliability of the safety barriers

input : $M_{S,i,j}$, $M_{F,i,j}$, $N_{S,i,j}$, $N_{F,i,j}$

output: $\left\{R^{(l)}, K^{(l)}\right\}_{l=1}^{N_l}$

Set initial values for $R^{(0)}, K^{(0)}, \pi^{(0)}$;

for $l = 1 : N_l$ **do**

  $R^{(l)} \leftarrow$ Generate a random sample from (10.17), where $K = K^{(l-1)}, \pi = \pi^{(l-1)}$;

  $K^* \leftarrow$ Generate a random sample from the proposal density, *i.e.*, Uniform(K_L, K_U);

  $p_{acc} \leftarrow$ Calculate p_{acc} using (10.19), where $R_{B,i,j} = R^{(l)}, \pi = \pi^{(l-1)}$;

  $r \leftarrow$ Generate a sample from Uniform$(0, 1)$;

  if $r \leq p_{acc}$ **then** $K^{(l)} \leftarrow K^*$;

  else $K^{(l)} \leftarrow K^{(l-1)}$;

  $\pi^{(l)} \leftarrow$ Generate a random sample from (10.7);

return $\left\{R^{(l)}, K^{(l)}\right\}_{l=1}^{N_l}$.

$$p_1(R_{B,i,j}, K) = \begin{cases} 0, \ K > K_U \text{ or } K < K_L, \\ \Delta \cdot R_{B,i,j}^{M_{S,i,j}+K\pi-1} \cdot (1 - R_{B,i,j})^{M_{F,i,j}+K(1-\pi)-1} \cdot \dfrac{1}{B\left(K\pi, K(1-\pi)\right)} \cdot \dfrac{1}{K_U - K_L}, \text{ otherwise.} \end{cases} \tag{10.20}$$

$$p\left(R_{B,i,j} \mid K, M_{S,i,j}, M_{F,i,j}\right) \propto R_{B,i,j}^{M_{S,i,j}+K\pi-1} \cdot (1 - R_{B,i,j})^{M_{F,i,j}+K(1-\pi)-1}. \tag{10.21}$$

10.2.5 A sequential Bayesian updating algorithm for DRA

Once the reliability of the safety barriers are updated, DRA can be done using Algorithm 10.4. The resulting $\left\{\mathbf{r}_C^{(l)}\right\}_{l=1}^{N_l - N_{burn-in}}$ approximate the posterior distribution of $\mathbf{r}_C$ updated at $t = t_j$. At each $t = t_j, j = 1, 2, \cdots, q$, Algorithm 10.4 is recursively applied for the DRA.

Algorithm 10.4: Sequential Bayesian updating for DRA (for $t = t_j$)

for $i = 1 : m$ **do**

  $\{N_{S,i,j}, N_{F,i,j}\} \leftarrow$ Generate pseudo-test data based on statistical failure data, using (10.10);

  $\left\{\mathbf{x}_{i,j}^{(k)}, w_{i,j}^{(k)}\right\}_{k=1}^{N_P} \leftarrow$ Particle filtering based on condition-monitoring data, using Algorithm 10.1;

  $\{M_{S,i,j}, M_{F,i,j}\} \leftarrow$ Generate pseudo-test data based on $\left\{\mathbf{x}_{i,j}^{(k)}, w_{i,j}^{(k)}\right\}_{k=1}^{N_P}$, using Algorithm 10.2;

  $\left\{R_{B,i,j}^{(l)}\right\}_{l=1}^{N_l - N_{burn-in}} \leftarrow$ Update $R_{B,i,j}$ using Algorithm 10.3;

for $l = N_{burn-in} + 1 : N_l$ **do**

  $R_{B_i} \leftarrow R_{B,i,j}^{(l)}, i = 1, 2, \cdots, n$;

  $\mathbf{r}_C^{(l-N_{burn-in})} \leftarrow$ Calculate the risk indexes using (10.1);

return $\left\{\mathbf{r}_C^{(l)}\right\}_{l=1}^{N_l - N_{burn-in}}$.

10.2.6 An application

In this section, we consider the High-Flow Safety System (HFSS) analyzed in [76] as a case study to demonstrate the application of the developed methods. The HFSS is a system installed on a hazardous material storage tank to

defend against potential accidents that might be caused by a High-Flow (HF) event, *i.e.*, the flow rate of the input pipeline becomes abnormally high for some reasons [76]. A defense-in-depth strategy is implemented by five safety barriers in the HFSS. We distinguish three categories of consequences (denoted by C_A, C_B and C_C, respectively), depending on their severity. A detailed descriptions of the safety barriers and the consequences can be found in [152].

Statistical data of occurrence of consequences C_A, C_B and C_C are available for 10 years from similar systems, while condition-monitoring data are available for two safety barriers subject to degradation of their built-in batteries. For a detailed presentation of the data, please refer to Sect. IV of [152].

DRA of the HFSS is made using Algorithm 10.4. A comparison is made between the results obtained by the developed method and those of the DRA method that only considers statistical failure data [76]. By using the method in [76], the reliability of safety barriers 1 and 3 are updated using only the statistical failure data. It can be seen from Fig. 10.2 that before $t = 7$ (years), the risk indexes estimated by both methods show the same trend but in this region, the risks estimated by the developed method is less severe than that estimated by the method in [76]. This is because, as shown in Fig. 2 of [152], the corresponding condition-monitoring data suggest that both the BPC and the HLA have high reliability in this region, since their safety margins are large compared to the uncertainties in their estimates. Having this information, we are more confident that the HFSS can reliably work to reduce the potential risk from an accident.

Significant differences between the two methods are observed when $t \geq 7$ (years) in Fig. 10.2: the developed method suggests that r_{C_A}, the conditional probability of normal operation, begins to decrease from $t > 7$ (years), while the method in [76] suggests that it remains relatively stable. Also, the credibility interval becomes significantly narrower than the one obtained from [76]. This can be explained by the differences in the posterior reliability values of safety barriers 1 and 3 given by the two methods: as shown in Fig. 5 of [152], if we only use the statistical failure data, $R_{B,1,t}$ is relatively stable over the entire range $[0, 10]$ (years); if we introduce the condition-monitoring data in Fig. 2 of [152], $R_{B,1,t}$ is stable from $t = 1$ to $t = 7$ (years), while it decreases dramatically when $t > 7$ (years). The same phenomenon can be observed on safety barrier 3, which results in the deviation of the two methods in Fig. 10.2. From the comparison, it is shown that by introducing condition-monitoring data, the developed method is able to capture the system-specific characteristics related to the degradation of the target system, and, therefore, provides a more informed description of the risk of the target system.

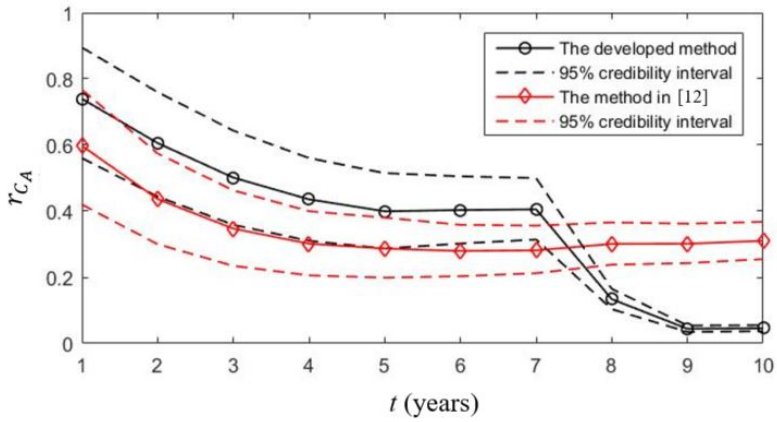


Figure 10.2: Comparison to the method in [76]

10.3 Fusing condition-monitoring data and inspection data for reliability assessment

In this section, we develop methods for fusing condition-monitoring data with inspection data for online reliability assessment. Condition monitoring data are online-collected by sensors and indirectly relate to component degradation; while inspection data are recorded in physical inspections that directly measure the component degradation. A hidden Markov Gaussian mixture model is developed in Sect. 10.3.1 for the condition-monitoring data. Modeling of inspection data and its integration with the condition-monitoring data are discussed in Sect. 10.3.2. Section 10.3.3 discusses reliability updating based on the integrated data sources. Finally, an application is presented in Sect. 10.3.4. The work in this section was previously published as a journal paper [141]. For more details, readers could refer to the original paper.

10.3.1 A Hidden Markov Gaussian Mixture Model for modeling condition monitoring data

In this section, we develop a HM-GMM to model condition monitoring data. An illustration of the model is given in Figure 10.3. It is assumed that the considered safety barrier degrades during its lifetime and the degradation process follows a discrete state discrete time Markov model $S(t)$ with a finite state space $S(t) \in \{S_1, S_2, \cdots, S_Q\}$, where $S(t)$ represents the health state of the safety barrier, Q is the number of health states, and $S\_1, S\_2, n\{\}cdots, S\_Q$ are in descending order of health (S_1 is the perfect functioning state, S_Q is the failure state). The evolution of the degradation process is characterized by the transition probability matrix of the Markov process, denoted by A where $A = \{a_{ij}\}$ and $a_{ij} = P(S(t_{k+1} = S_j) \mid S(t_k) = S_i)$, $k = 1, 2, \cdots, q$, $1 \leq i, j \leq Q$. The initial state distribution of the Markov process is denoted by $\boldsymbol{\pi} = [\pi_1, \pi_2, \cdots, \pi_Q]$ where $\pi_i = P(S(t_0) = S_i), 1 \leq i \leq Q$. It should be noted that repairs are not considered in this paper to simplify the calculation. Therefore, $S(t)$ can only transit to a worse state

and cannot move backwards. Besides, the failure state S_Q is an absorbing state. However, this model can be easily extended to repairable component: only the transition matrix needs to be modified to allow backward jumps, which represent the repair of the safety barrier. The developed algorithms, can, then, be extended naturally.

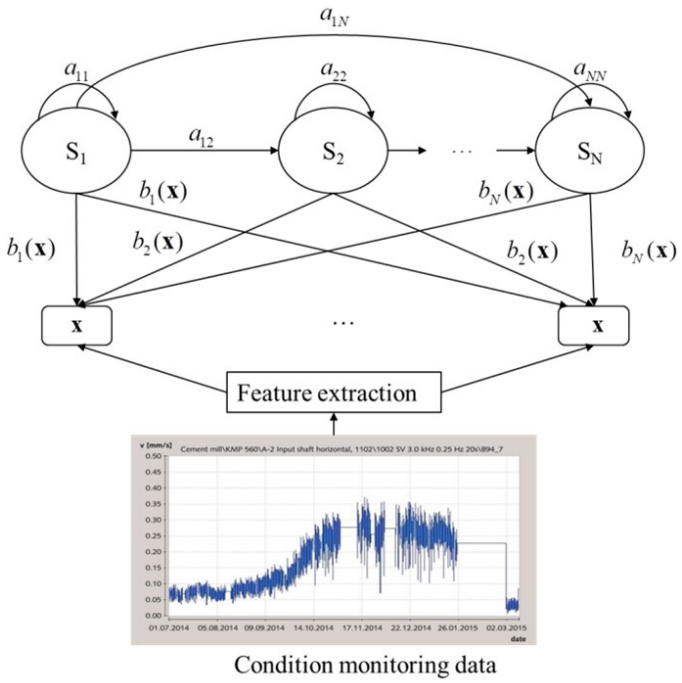


Figure 10.3: Description of the HM-GMM.

Condition monitoring data $\mathbf{c}(t)$ are available at $t = t_k, k = 1, 2, \cdots, q$. In practice, $\mathbf{c}(t)$ contains only raw signals, which cannot be directly used for degradation modeling and analysis. Feature extraction, as shown in Figure 10.3, is needed to extract degradation features from $\mathbf{c}(t)$. For example, vibration signals are usually used as condition monitoring data for bearings [67]. The raw vibration signals, however, need to be preprocessed to extract features for degradation characterization. The commonly used degradation features include entropy, root mean square (RMS), kurtosis, etc [67]. In this paper, we refer to these extracted features as degradation indicators and denote them by $\mathbf{x}(t)$, where $\mathbf{x}(t) = \left[x_1(t), x_2(t), \cdots, x_{n_{feature}}(t)\right]$ and $n_{feature}$ is the number of the degradation features.

As the safety barrier degrades, the degradation indicator $\mathbf{x}(t)$ exhibits distinct patterns. To capture such patterns and the uncertainty associated with them, it is assumed that at each degradation state $S_i, 1 \leq i \leq Q$, the values of the degradation indicators $\mathbf{x}$ follow a multivariate Gaussian distribution $b_i(\mathbf{x}) = p(\mathbf{x}\,|S(t) = S_i\,) = N(\mathbf{x}\,|\mu^{(i)}, \mathbf{\Sigma}^{(i)}), i = 1, 2, \cdots, Q$, as shown in Figure 3. The mean values vector $\mu^{(i)}$ captures the degradation pattern at each degradation state, while the covariance matrix $\mathbf{\Sigma}^{(i)}$ captures the uncertainty in the condition monitoring data. An overall picture of the HM-GMM is given in Figure 10.3. Conceptually, we denote the HM-GMM compactly as λ=$\{\pi, A, \mu, \mathbf{\Sigma}\}$, where π is the initial state distribution, A is the transition probability matrix, μ= $[\mu_1, \mu_2 \cdots, \mu_Q]$ is a vector of the mean values and $\mathbf{\Sigma} = \left[\Sigma^{(1)}, \Sigma^{(2)}, \cdots, \Sigma^{(Q)}\right]$ is a collection of the covariance matrices of the multivariate Gaussian distribution, respectively.

The HM-GMM model can be trained in an offline phase based on an expectation maximization algorithm. The trained model can, then, be used in an online phase that allows estimating degradation states $S(t)$ based on the collected condition-monitoring data. For details of the training and estimation algorithms, please refer to our publication in [141]. Hereafter, we denote the estimated degradation state from the condition-monitoring data as $S_{CM}(t)$.

10.3.2 Integrating condition monitoring data with inspection data

To update and predict the reliability, one needs to estimate the degradation state first. Let S_{IN} denote the degradation state estimated from inspection data and S denote the true degradation state. In practice, S_{IN} is subject to uncertainty due to potential imprecision in the inspection and recording by the maintenance personnel. To model such uncertainty, in this paper, we assume that the reliability of inspection is R_{IN}, and that the maintenance personnel correctly identify the true degradation state with a probability R_{IN}, whereas an inspection error can occur with probability $(1 - R_{IN})$. When an inspection error occurs, it is further assumed that the probabilities for each of the possible degradation states being erroneously identified as the true degradation state are equal to each other:

$$P(S_{IN} = S_i \,|S\,) = \begin{cases} R_{IN}, & S = S_i \\ \frac{1-R_{IN}}{Q-1}, & S \neq S_i\,, \end{cases} \tag{10.22}$$

where Q is the number of degradation states. It is should be noted that other inspection models might also be assumed, depending on the actual problem setting.

In this paper, a BN is developed to describe the dependencies among S, S_{IN}, S_{CM}, as shown in Figure 5. The BN in Figure 10.4 is constructed based on the assumption that given the true degradation state S, the estimated degradation state from condition monitoring data and inspection data are conditional-independent.

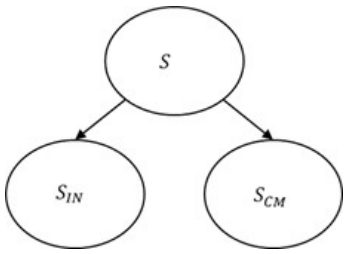


Figure 10.4: A BN model for data integration.

Based on the BN in Figure 5, we have

$$P\,(S, S_{IN}, S_{CM}) = P\,(S_{IN}\,|S)\,P\,(S_{CM}\,|S)\,P\,(S)\,. \tag{10.23}$$

In (10.23), $P\,(S)$ measures the prior belief of the analysts on the current degradation states. We assume that $P\,(S)$

is a uniform distribution over all the possible degradation states, indicating that there is no further information to distinguish the states.

The conditional probability distribution $P(S_{IN}|S)$ describes the uncertainty in the inspections and is derived based on (10.22). In (10.22), the reliability of the inspection can be estimated from historical data or assigned based on expert judgments. The conditional probability distribution $P(S_{CM}|S)$ measures the trust one has on the estimated degradation state based on condition monitoring data. Its values can be estimated from validation test data. However, in practice, as validation tests are not always available, $P(S_{CM}|S)$ might also be assigned by experts considering the measurement uncertainty of the sensors and the distance between the neighboring degradation states.

Once the condition monitoring data and inspection data are available, the observed values of S_{IN} and S_{CM} are known. Suppose we have $S_{CM} = S_j$ and $S_{IN} = S_i$. It should be noted that we choose the state with maximal posterior probability as the observation value of S_{CM}. The two data sources can be naturally integrated by calculating the posterior distribution of S given the two data sources, denoted by $P_{INT}(S)$. Based on the BN in Figure 10.4, we have:

$$\begin{aligned} P_{INT}(S) &\triangleq P(S|S_{IN} = S_i, S_{CM} = S_j) \\ &= \frac{P(S, S_{IN} = S_i, S_{CM} = S_j)}{P(S_{IN} = S_i, S_{CM} = S_j)} \\ &= \frac{P(S_{IN} = S_i|S)\,P(S_{CM} = S_j|S)\,P(S)}{P(S_{IN} = S_i, S_{CM} = S_j)} \end{aligned} \tag{10.24}$$

10.3.3 Reliability updating and prediction

Given the estimated posterior distribution in (10.24), the reliability of the safety barrier can be updated. Suppose the current time is t_k, the updated reliability can be calculated by:

$$R_{SB}(t_k) = \sum\nolimits_{S \in W} \mathsf{P}_{INT,t_k}(S), \tag{10.25}$$

where W is the working set that contains all the working states; $\mathsf{P}_{INT,t_k}(S)$ is the posterior probability of the true degradation state after integrating the two data sources at $t = t_k$ and is calculated from (10.24).

Furthermore, at $t = t_k$, we can also predict the reliability of the safety barriers at a future time t_{Fut}. For this, the distribution of the degradation states at $t = t_{Fut}$ is predicted first, using Chapman-Kolmogorov equation [133] and the trained model from the offline step:

$$P_{INT,t_{Fut}}(S) = P_{INT,t_k}(S) \times \widehat{A}^{(t_{Fut}-t_k)}.$$

The reliability at $t = t_k$, can be predicted as:

$$R_{SB}(t_{Fut}) = \sum\nolimits_{S \in W} P_{INT,t_{Fut}}(S).$$

10.3.4 An application

In this section, the developed method is applied for dynamic risk assessment of an Anticipated Transient Without Scram (ATWS) accident of a NPP [61]. In this original ETA of the ATWS, the failure probabilities of the safety barriers are assumed to be constant values. In practice, however, these probabilities might change due to various degradation mechanisms. Take the recirculation pump as an example. According to [83], most field failures of the recirculation pump are caused by the degradation of the bearing inside the pump, which makes the failure probability of the recirculation pump time-dependent. Therefore, the bearing of the recirculation pump is assumed to be degrading and its condition-monitoring data bearing come from the bearing degradation dataset from university of Cincinnati [123]. Inspections are conducted at three time instants, i.e., $t = 30, 35, 50$ (d), respectively. The inspection data at the three time instants are given in Table 10.1. In Table 10.1, we also show the true degradation states (S) and the estimated degradation states using condition-monitoring data (S_{CM}). More detailed description of the case study can be found in our paper [141].

Table 10.1: Values of inspection data at different time instants.

	$t = 30$ (d)	$t = 35$ (d)	$t = 50$ (d)
S	S_2	S_3	S_3
S_{CM}	S_2	S_2	S_3
S_{IN}	S_2	S_3	S_2

The results of risk updating and prediction are given in Figure 10.5. In Figure 10.5, we also show the results from using only condition monitoring data and inspection data, for comparison. As shown in Figure 15(a), at $t = 30$ (d), the results from all the three methods are close to each other. This can be explained from Table 10.1: at $t = 30$ (d) both data sources correctly identify the true degradation states. However, when compared to the true risk values, the updated and predicted risks from all the three methods show relatively large discrepancies. This discrepancy is mainly due to the estimation errors in the offline step, as we have only four samples in the training data set. A possible way to increase the accuracy of risk updating is, then, to increase the sample size of the training data in the offline step.

It can be seen from Table 10.1 that at $t = 35$ (d) the inspection data give correct information on the current degradation state while condition monitoring data do not. From Figure 10.5(b), it can be seen that the developed data-integration method improves the DRA results from the condition monitoring data-based method, as it integrates

the correct information from inspection data. On the other hand, when the inspection data fail to give the correct information, it can be seen from Figure 10.5(c) that the developed data integration method can also correct the misleading results obtained from using only the inspection data. Hence, in general, applying the developed data integration method can achieve a more robust DRA result than using the two data sources individually.

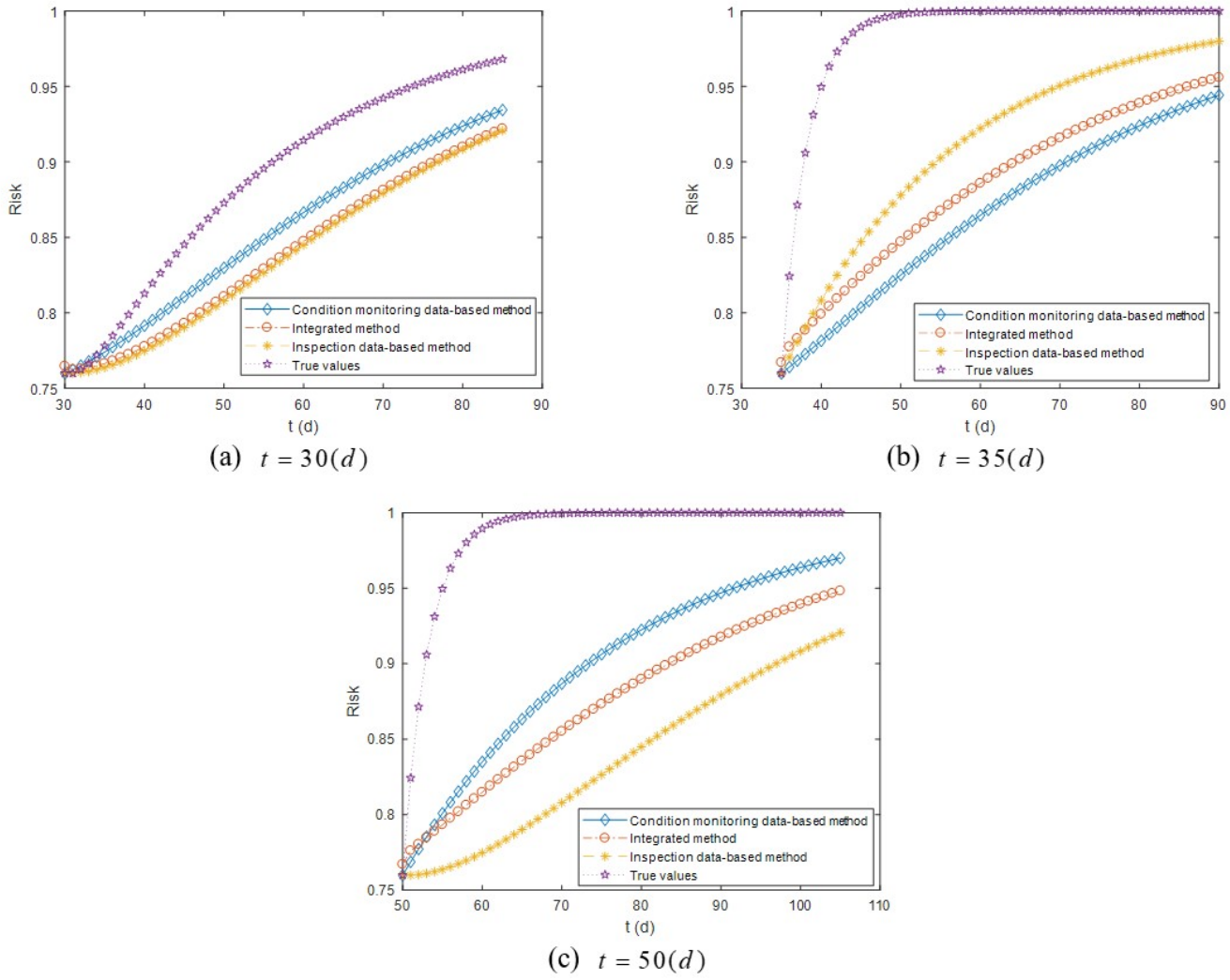


Figure 10.5: The results of risk updating and prediction.

In Figure 10.6, we compare the developed DRA method with the conventional ETA method in [61]. It can be seen from Figure 10.6 that the results from the developed DRA method are closer to the true risk values than those of the standard ETA. This is because through the integration of inspection and conditon monitoring data, the developed method is able to capture the time-dependent behavior of the recirculation pump resulting from the degradation of the bearing. The standard ETA, however, fails to capture such time-dependencies as it assumes that the event probabilities do not change although the real system/component ages over time.

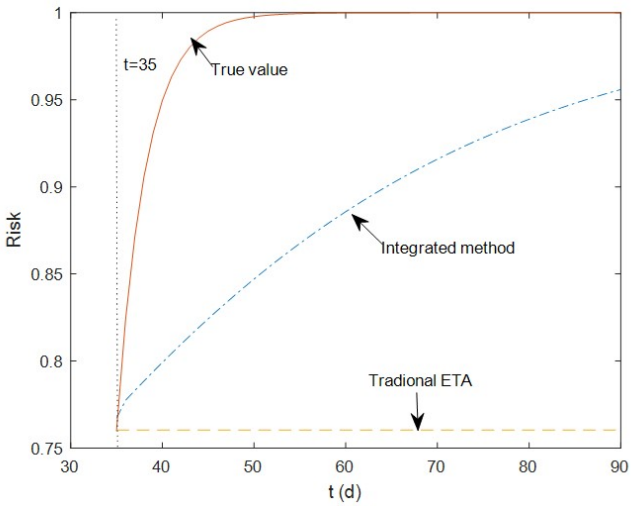


Figure 10.6: Comparisons of to traditional ETA (at $t = 35$ (d)).

10.4 Fusing expert knowledge with condition-monitoring data for RUL prediction

In this section, we develop a mixture of Gaussians-evidential hidden Markov model (MoG-EHMM) to fuse expert knowledge with CM information. The MoG-EHMM is formally defined as a three-layer model (See details in Sect. 10.4.1. The MoG-EHMM-based RUL prediction comprises of two phases: offline and online. Both phases can elicit Expert knowledge. In the offline phase, training data are collected from a population of similar systems. Then, health indicators (HIs) are extracted from the original training data through feature extraction. Evidential Expectation-Maximization (E^2M) algorithm is implemented to estimate the parameters of MoG-EHMM for model training (See details in Sect. 10.4.2). In the online phase, CM information is collected from a new system. Based on the extracted HIs and the online expert knowledge, forward algorithm [115] is exploited recursively for health state inference, system reliability updating, and the RUL prediction. (See details in Sects. 10.4.3, and 10.4.4). This work was previously published as a journal paper [139]. More information could be found in the original publication.

10.4.1 Model Formulation

The MoG-EHMM comprises of three-layers: true degradation layer, observation layer, and knowledge layer, as shown in Fig. 10.7. The true degradation layer models the true (but unobservable) degradation process. It is partially hidden because some knowledge of the health state of a system is available from experts. The observation layer represents the HIs extracted from signals, and the knowledge layer quantifies the expert knowledge by the contour functions under the BFT.

In the true degradation layer, it is assumed that the degradation of a system is multi-state and can be modeled by a discrete- time discrete-state Markov model. Let $S(t)$ denote the system states associated with degradation,

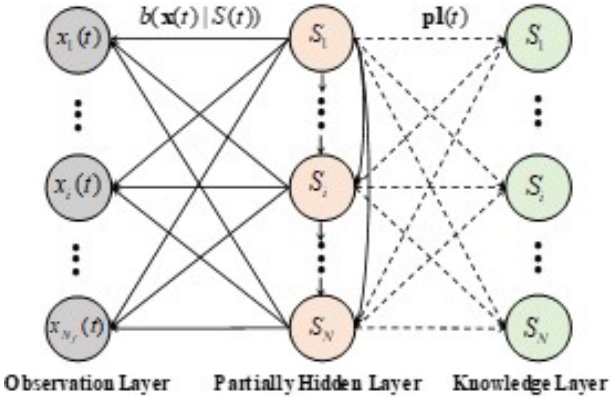


Figure 10.7: The proposed MoG-EHMM.

where $S(t) \in S_1, S_2, ..., S_N$ and *N* is the number of system states. The number of states can be determined by expert experience [131]. Let $S_1, S_2, ..., S_N$ be in descending order of performance levels where S_1 is the perfect functioning state and S_N is the complete failure state. The one- step transition probability from state S_i to S_j is denoted as$a_{ij} = \Pr\{S(t+1) = S_j | S(t) = S_i\}$ $(t = 1, 2, ..., T, 1 \leq i, j \leq N)$. The corresponding transition probability matrix is denoted by $\mathbf{A} = [a_{ij}]_{N \times N}$, where $\sum_{j=1}^{N} a_{ij} = 1$. Note that maintenance action is not considered in the present study, i.e.,$S(t)$ can only transit to worse states. The resulting hidden Markov model is the so-called left-right or Bakis model in HMMs [115]. The initial state probability distribution is represented by $\pi = [\pi_1, \pi_2, ..., \pi_i, ..., \pi_N]$ where $\pi_i = \Pr\{S(0) = S_i\}$ $(1 \leq i \leq N)$. System reliability$S(t) \in [S_1, S_2, ..., S_i, ..., S_N]$ is defined as the probability that the performance level of the system is not lower than a threshold state S_F,

$$R_S(t) = \sum\nolimits_{S_i \leq S_F} \Pr\{S(t) = S_i\} \tag{10.26}$$

In the observation layer, continuous CM information denoted by $\mathbf{c}(t)$, such as vibration signals and acceleration signals, can be collected from sensors. Through feature extraction, HIs can be reconstructed from $\mathbf{c}(t)$. Let $\mathbf{x}(t)$ $(t = 1, 2, ..., T$) denote the extracted HIs, where $\mathbf{x}(t) = [x_1(t), x_2(t), ..., x_{N_f}(t)]$ and N_f is the number of HIs. Examples of the HIs include root mean square, mean value, and kurtosis. A Gaussian mixture model (GMM) is utilized to characterize the continuous evolution behaviors of the HIs and the uncertainty associated with them. In the GMM, the emission probability $b(\mathbf{x}(t)|S(t))$ represents the probability of observing the current values of HIs $\mathbf{x}(t)$, given that the current state is $S(t)$, i.e., $b_i(\mathbf{x}(t)|S(t)) = \Pr\{\mathbf{x}(t)|S(t) = S_i\} = N(\mathbf{x}(t)|\mu_i, \mathbf{\Sigma}_i)$, where μ_i is the mean value of $\mathbf{x}(t)$ under the given hidden state S_i, while the covariance matrix $\mathbf{\Sigma}_i$ captures the uncertainty associated with $\mathbf{x}(t)$. Note that the number of Gaussian components is determined in this paper by minimizing the Akaike information criterion (AIC) [131].

In the knowledge layer, expert knowledge on the health state of a system is given in the form of mass function $\mathbf{m}(t) = [m_{S_1}(t), m_{S_2}(t), ..., m_{S_N}(t), ..., m_\Omega(t)]$ or contour function $\mathbf{pl}(t) = [pl_{S_1}(t), pl_{S_2}(t), ..., pl_{S_N}(t)]$ $(t = 1, 2, ..., T$)

under the BFT. A commonly-used format for experts to express their knowledge in terms of contour function [118]:

$$pl_{S_i}(t) = \begin{cases} 1, \text{ if} S_{Expert} = S_i, \\ \rho, \text{ otherwise.} \end{cases} \tag{10.27}$$

where $S_{Expert}(t)$ is the current health state judged by experts, ρ ($0 \leq \rho \leq 1$) is the non-specificity coefficient that quantifies the epistemic uncertainty in the expert judgments: $\rho = 0$ indicates precise knowledge elicited by experts, whereas $\rho = 1$ corresponds to the non-informative knowledge. A larger value of ρ means experts are more uncertain in his/her judgments. Note that $S_{Expert}(t)$ is not guaranteed to be coincide with the true health state, denoted by $S_{True}(t)$, as the experts might sometimes provide incorrect judgments.

Expert knowledge can also be elicited in terms of mass function. In fact, as shown in Proposition 1 in the Appendix A in the supplementary file, mass function is equivalent to contour function in terms of eliciting knowledge from the experts. If the mass function is elicited, it can be converted into the contour functions before fusing with CM information. Therefore, we only present the developed methods in terms of contour functions in this work.

10.4.2 Parameter Estimation of MoG-EHMM in the Offline Phase

In the offline phase, the training data, denoted as $\mathbf{c}_{Tr}^{(k)}(t)$ ($k = 1, 2, ..., K$, $t = 1, 2, ..., T$) need to be collected from a population of K identical systems. The number of training data, i.e., *K*, should be as large as possible, because the size of the training dataset could directly impact the training performance. Through feature extraction, the HIs $\mathbf{x}_{Tr}^{(k)}(t)$ ($k = 1, 2, ..., K$, $t = 1, 2, ..., T$) can be extracted. Based on the training data $\mathbf{x}_{Tr}^{(k)}(t)$ and its corresponding expert contour function $\mathbf{pl}_{Tr}^{(k)}(t)$ ($k = 1, 2, ..., K$, $t = 1, 2, ..., T$), the parameter of the MoG-EHMM $\hat{\theta} = (\hat{\pi}, \hat{\mathbf{A}}, \hat{\mu}, \hat{\mathbf{\Sigma}})$ can be estimated by maximizing the likelihood of observing $\mathbf{x}_{Tr}^{(k)}(t)$ and $\mathbf{pl}_{Tr}^{(k)}(t)$, ($k = 1, 2, ..., K$, $t = 1, 2, ..., T$)

$$\begin{aligned} \hat{\theta} &= \arg \max_{\theta} \ L(\mathbf{x}, \mathbf{pl}|\theta) = \Pr\{\mathbf{x}_{Tr}^{(1)}(t), ..., \mathbf{x}_{Tr}^{(K)}(t), \mathbf{pl}_{Tr}^{(1)}(t), ..., \mathbf{pl}_{Tr}^{(K)}(t)|\theta\} \\ &= \arg \max_{\theta} \ L(\mathbf{x}, \mathbf{pl}|\theta) = b(\mathbf{x}_{Tr}^{(1)}(t), ..., \mathbf{x}_{Tr}^{(K)}(t)|\theta) \oplus (\mathbf{pl}_{Tr}^{(1)}(t), ..., \mathbf{pl}_{Tr}^{(K)}(t)) \\ &= \arg \max_{\theta} \ L(\mathbf{x}, \mathbf{pl}|\theta) = \prod_{k=1}^{K} \prod_{t=1}^{T} b(\mathbf{x}_{Tr}^{(k)}(t)|\theta) \oplus \mathbf{pl}_{Tr}^{(k)}(t) \end{aligned} \tag{10.28}$$

where $b(\mathbf{x}_{Tr}^{(k)}(t)|\theta)$ is the emission probability of the *k*th training data at time *t* given the parameters θ of the MoG-EHMM. Note that the result of $b(\mathbf{x}_{Tr}^{(k)}(t)|\theta) \oplus \mathbf{pl}_{Tr}^{(k)}(t)$ is still a probability measure as $b(\mathbf{x}_{Tr}^{(k)}(t)|\theta)$ can be regarded as a Bayesian mass [118]. Directly resolving (10.28) is challenging because the true states are partially hidden. The E^{2}M algorithm can be implemented to calculate $L(\mathbf{x}, \mathbf{pl}|\theta)$ iteratively:

***E*-Step**: Compute the expectation of $L(\mathbf{x}, \mathbf{pl}|\hat{\theta})$ given the current estimates $\hat{\theta}$.

***M*-Step:** Maximize the log-likelihood function obtained in the *E*-Step and calculate a new maximum likelihood

estimate for the unknown parameters $\hat{\theta}$.

To implement the E^{2}M algorithm, two auxiliary variables, namely the forward variable $\alpha_{S_j}^{(k)}(t)$ and backward variable $\beta_{S_j}^{(k)}(t)$ are introduced. In this work, $\alpha_{S_j}^{(k)}(t)$ is defined as the probability of observing $\mathbf{x}_{Tr}^{(k)}(1), ..., \mathbf{x}_{Tr}^{(k)}(t)$ and $\mathbf{pl}_{Tr}^{(k)}(1), ..., \mathbf{pl}_{Tr}^{(k)}(t)$ with the current state S_j given the parameter θ:

$$\alpha_{S_j}^{(k)}(t) = \left(b(X_{Tr}^{(k)}(t) \mid \theta) \oplus \mathbf{pl}_{Tr}^{(k)}(t)\right)[S_j] \tag{10.29}$$

for $1 \leq j \leq N$ ($k = 1, 2, ..., K$,$t = 1, 2, ..., T$). It can be verified that

$$\begin{cases} \alpha_{S_j}^{(k)}(1) = \pi_i \times \left(b(\mathbf{x}_{Tr}^{(k)}(1)|\theta) \oplus \mathbf{pl}_{Tr}^{(k)}(1)\right)[S_j] \\ \alpha_{S_j}^{(k)}(t+1) = \left(b(\mathbf{x}_{Tr}^{(k)}(t+1)|\theta) \oplus \mathbf{pl}_{Tr}^{(k)}(t+1)\right)[S_j] \times \left[\sum_{i=1}^{N} \alpha_{S_i}^{(k)}(t) a_{ij}\right] \end{cases} \tag{10.30}$$

The backward variable $\beta_{S_j}^{(k)}(t)$ is defined as the probability of observing $\mathbf{x}_{Tr}^{(k)}(t+1), ..., \mathbf{x}_{Tr}^{(k)}(T)$ and $\mathbf{pl}_{Tr}^{(k)}(t+1), ..., \mathbf{pl}_{Tr}^{(k)}(T)$ ($k = 1, 2, ..., K$, $t = 1, 2, ..., T-1$) given the current state S_j and the parameter θ

$$\begin{aligned} \beta_{S_j}^{(k)}(t) &= \Pr\{\mathbf{x}_{Tr}^{(k)}(t+1), \mathbf{pl}_{Tr}^{(k)}(t+1) | S(t) = S_j, \theta\} \\ &= b_j(\mathbf{x}_{Tr}^{(k)}(t+1)|\theta) \oplus \mathbf{pl}_{Tr}^{(k)}(t+1) \end{aligned} \tag{10.31}$$

for $1 \leq i, j \leq N$ ($k = 1, 2, ..., K$, $t = 1, 2, ..., T-1$). It can also be verified that

$$\begin{cases} \beta_{S_j}^{(k)}(T) = 1 \quad \text{for } 1 \leq j \leq N \\ \beta_{S_j}^{(k)}(t+1) = \left[\sum_{i=1}^{N} \left(b(\mathbf{x}_{Tr}^{(k)}(t+1)|\theta) \oplus \mathbf{pl}_{Tr}^{(k)}(t+1)\right)[S_j] \times a_{ij}\right] \beta_{S_j}^{(k)}(t) \end{cases} \tag{10.32}$$

The probability of being in state S_j at time *t* given $\mathbf{x}_{Tr}^{(k)}(1), ..., \mathbf{x}_{Tr}^{(k)}(t)$, $\mathbf{pl}_{Tr}^{(k)}(1), ..., \mathbf{pl}_{Tr}^{(k)}(t)$ ($k = 1, 2, ..., K$,$t = 1, 2, ..., T$) and the parameter θ, denoted as $\gamma_{S_j}^{(k)}(t)$, can be calculated by

$$\gamma_{S_j}^{(k)}(t) = \frac{\alpha_{S_j}^{(k)}(t)\beta_{S_j}^{(k)}(t)}{\sum_{j=1}^{N} \alpha_{S_j}^{(k)}(t)\beta_{S_j}^{(k)}(t)} \tag{10.33}$$

and the probability of the *k*th training data being in state S_i at time *t* while in state S_j at time *t*+1, denoted as $\xi_{S_i,S_j}^{(k)}(t)$, can be computed by

$$\xi_{S_i,S_j}^{(k)}(t) = \frac{\alpha_{S_i}^{(k)}(t) a_{ij} \left(b(\mathbf{x}_{Tr}^{(k)}(t+1)|\theta) \oplus \mathbf{pl}_{Tr}^{(k)}(t+1)\right)[S_j]\beta_{S_j}^{(k)}(t+1)}{\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_{S_i}^{(k)}(t) a_{ij} \left(b(\mathbf{x}_{Tr}^{(k)}(t+1)|\theta) \oplus \mathbf{pl}_{Tr}^{(k)}(t+1)\right)[S_j]\beta_{S_j}^{(k)}(t+1)} \tag{10.34}$$

After calculating $\xi_{S_i,S_j}^{(k)}(t)$ and $\gamma_{S_j}^{(k)}(t)$ for all training data, the estimate of the initial state probability $\hat{\pi}_i$ ($1 \leq i \leq N$)

can be calculated by

$$\hat{\pi}_i = \frac{\sum_{k=1}^{K} \gamma_{S_j}^{(k)}(1)}{K} \tag{10.35}$$

The estimate of the one-step transition probability a_{ij} ($1 \leq i, j \leq N$) is

$$\hat{a}_{ij} = \frac{\sum_{k=1}^{K} \sum_{t=1}^{T} \xi_{S_i,S_j}^{(k)}(t)}{\sum_{k=1}^{K} \sum_{t=1}^{T} \gamma_{S_i}^{(k)}(t)} \tag{10.36}$$

The estimates of the mean value vector and the covariance matrices of the MoGs can be calculated by

$$\hat{\mu}_i = \frac{\sum_{k=1}^{K} \sum_{t=1}^{T} \gamma_{S_i}^{(k)}(t) \mathbf{x}_{Tr}^{(k)}(t)}{\sum_{k=1}^{K} \sum_{t=1}^{T} \gamma_{S_i}^{(k)}(t)} \tag{10.37}$$

and

$$\hat{\mathbf{\Sigma}}_i = \frac{\sum_{k=1}^{K} \sum_{t=1}^{T} \gamma_{S_i}^{(k)}(t) (\mathbf{x}_{Tr}^{(k)}(t) - \hat{\mu}_i)(\mathbf{x}_{Tr}^{(k)}(t) - \hat{\mu}_i)'}{\sum_{k=1}^{K} \sum_{t=1}^{T} \gamma_{S_i}^{(k)}(t)}, \tag{10.38}$$

respectively. The MoG-EHMM training procedure is summarized in Algorithm 5. The initial value for μ_0 can be set by the *K*-means clustering algorithm, while $\pi_0, \mathbf{A}_0, \mathbf{\Sigma}_0$ can be initialized by assuming non-informative knowledge. Convergence of Algorithm 5 is checked by comparing the relative deviation of the maximum log-likelihood between two adjacent iterations to a pre-specified threshold ε, say ε=10^{-6} as used in this work. Note that the forward and backward variables should be normalized in each step to avoid exponentially converging to zero.

Algorithm 10.5: Parameter estimation of MoG-EHMM $\hat{\theta}$

Require: Initial values of $\hat{\theta}$, denoted as θ_0;
Training data $\mathbf{x}_{Tr}^{(k)}(1), ..., \mathbf{x}_{Tr}^{(k)}(T)$;
Expert knowledge $\mathbf{pl}_{Tr}^{(k)}(1), ..., \mathbf{pl}_{Tr}^{(k)}(T)$.
Output: The estimated parameters $\hat{\theta}$ of the MoG-EHMM.
1: Set $\theta^{(q)} = \theta_0$; *q*=1;
2: **For** *k*=1 to *K* **do**
3: **For** *t*=1 to *T* **do**
4: Calculate $\alpha_{S_j}^{(k)}(t)$, $\beta_{S_j}^{(k)}(t)$, $\gamma_{S_j}^{(k)}(t)$, $\xi_{S_i,S_j}^{(k)}(t)$ by (7),(9)-(11);
5: Normalize $\alpha_{S_j}^{(k)}(t)$ and $\beta_{S_j}^{(k)}(t)$;
6: **End For**;
7: **End For**;
8: Calculate the parameter $\hat{\theta}$=$\theta^{(q+1)}$ by (12)-(15);
9: **If** , $\hat{\theta}$=$\theta^{(q+1)}$
10:$\hat{\theta}$=$\theta^{(q+1)}$; **Break**;
11: **Else** $\theta = \theta^{(q+1)}$; Go to Step 2;
12: **End If**.

10.4.3 Health State Inference and Reliability Updating in the Online Phase

In the online phase, the CM information, denoted as $\mathbf{c}_{CM}(t)$ ($t = 1, 2, ..., T$) is collected from a particular individual system of interest. Similar to the offline phase, once $\mathbf{c}_{CM}(t)$ is collected, the HIs $\mathbf{x}_{CM}(t)$ ($t = 1, 2, ..., T$) are

extracted through feature extraction. Expert knowledge can also be elicited in terms of contour functions for the states, denoted by $\mathbf{pl}_{CM}(t_k)$. The two information sources can be merged to estimate the true health state of the system and update reliability estimation based on the trained MoG-EHMM in the offline phase. Let $\mathbf{p}_{CM}(t_k) = [p_{CM,S_1}(t_k), p_{CM,S_2}(t_k), ..., p_{CM,S_i}(t_k), ..., p_{CM,S_N}(t_k)]$ represents the posterior state probability distribution of the new system updated by $\mathbf{x}_{CM}(t_k)$ and $\mathbf{pl}_{CM}(t_k)$ up to time t_k, that is

$$p_{CM,S_i}(t_k) = \Pr\{S(t_k) = S_i | \mathbf{x}_{CM}(t_k), \mathbf{pl}_{CM}(t_k), \hat{\theta}\}. \tag{10.39}$$

From Bayesian theorem, $p_{CM,S_i}(t_k)$ can be readily computed based on the forward variable using $\mathbf{x}_{CM}(t_k)$ and $\mathbf{pl}_{CM}(t_k)$, and $\hat{\theta}$ as following

$$p_{CM,S_i}(t_k) = \frac{\Pr\{S(t_k) = S_i, \mathbf{x}_{CM}(t_k), \mathbf{pl}_{CM}(t_k) | \hat{\theta}\}}{\Pr\{\mathbf{x}_{CM}(t_k), \mathbf{pl}_{CM}(t_k) | \hat{\theta}\}} = \frac{\alpha_{S_i}(t_k)}{\sum\limits_{S_i} \alpha_{S_i}(t_k)} \tag{10.40}$$

Let $S_{MAP}(t_k)$ be the most likely state at time t_k. It can be determined by maximizing the posterior probability $p_{CM,S_i}(t_k)$ (i=1,2,...,N)

$$S_{MAP}(t_k) = \underset{i=1,2,...,N}{\arg} \max\ p_{CM,S_i}(t_k), \quad \forall t_k = 1, 2, ..., T. \tag{10.41}$$

Similarly, the system reliability can be updated by the posterior probability distribution $\mathbf{p}_{CM}(t_k)$ and the transition probability matrix $\hat{\mathbf{A}}$ estimated in the offline phase.

$$R_S(t') = \sum_{S_i \leq S_F} \mathbf{p}_{CM}(t_k) \times \hat{\mathbf{A}}^{(t')}, \tag{10.42}$$

where t' is the time elapsed after the running time t_k of the specific new system.

10.4.4 RUL Prediction

Given the failure threshold state S_F , if $S_{MAP}(t_k) > S_F$, the RUL of the system is definitely zero. Otherwise, let $\bar{\tau}$ denotes the first passage time to the failure state $\{S_j\}$ where $S_j > S_F$

$$\bar{\tau} = \inf\{\tau : \ \ S(\tau) > S_F | S_{MAP}(t_k) \leq S_F\}. \tag{10.43}$$

Hence, the RUL of the system, i.e., t', is

$$t' = \bar{\tau} - t_k, \tag{10.44}$$

and the probability mass function of t' can be calculated by

$$q_{t'=t} = \Pr\{t' = t\} = \Pr\{\bar{\tau} = t_k + t\}, \quad t = 1, 2, ... \tag{10.45}$$

Based on the total probability law, the probability mass function of RUL t' can be computed recursively

$$\begin{aligned} q_{t'=t} = & \sum_{S_j > S_F} \Pr\{S(t_k + t) = S_j | S(t_k) = S_i\} \\ & - \sum_{l=1}^{t-1} \sum_{S_j > S_F} \Pr\{S(t_k + t) = S_j | t' = l\} \times q_{t'=l} \end{aligned}, \tag{10.46}$$

where $\sum_{S_j > S_F} \Pr\{S(t_k + t) = S_j | S(t_k) = S_i\} = 1 - R_S(t)$. In general, it is difficult to calculate $\Pr\{S(t_k + t) = S_j | t' = l\}$ for $1 \leq l \leq t-1$, because the recovery from the failure state to a functioning state is possible for repairable systems. In this paper, as we consider only non-repairable systems, (10.46) reduces to

$$q_{t'=t} = R_S(t-1) - R_S(t). \tag{10.47}$$

The developed methods were applied on a simulation case study and a real-world case study. Details of the application can be found in our paper [139]. The results showed that by introducing the expert knowledge, the performances of reliability assessment, health state inference and RUL prediction can be substantially improved.

10.5 Summary of major contributions

In this chapter, we presented our major works related to fusing multiple heterogeneous data sources for online risk and reliability assessment. Focusing on the research question identified in Sect. 10.1, the main contributions of our works can be summarized as follows:

1. A hierarchical Bayesian model and a Bayesian updating algorithm, which integrates Particle Filtering (PF) with Markov Chain Monte Carlo (MCMC), are developed to integrate the statistical failure data and condition-monitoring data for online reliability assessment. A comparison to the traditional method which only uses statistical failure data shows that by introducing condition-monitoring data on the system degradation process, it is possible to capture the system-specific characteristics, and, therefore, provide a more complete and accurate description of the risk of the target system.

2. A framework based on hidden Markov Gaussian mixture models and Bayesian networks is developed to fuse condition-monitoring and inspection data for dynamic risk assessment. The application results show that integrating the two data sources into the DRA gives more accurate and robust results than using any one of

the two individual data sources.

3. A mixture of Gaussians-evidential hidden Markov model is put forth for RUL prediction by fusing expert knowledge and CM information under the belief function theory framework. Simulation results and real case study showed that by introducing the expert knowledge, the performances of reliability assessment, health state inference and RUL prediction can be substantially improved.

Chapter 11

FUTURE RESEARCH PLANS

11.1 Scientific projects after HDR - Smart reliability engineering: Facing the challenges and opportunities of industry 4.0

Nowadays, industries are swiftly shifting towards the direction of industry 4.0, an initiative that intends to make the industries smart through the intelligent networking of machines and industrial processes with the help of information and communication technology [45]. To support industry 4.0, systems distributed in different places need to be connected through the Internet, creating a large-scale cyber-physical systems-of-systems [94]. This trend will create significant challenges to the reliability of the systems. First, the system's scale makes it an extremely complex system, bringing many potential reliability issues. For example, unexpected failures might emerge at the system level. Besides, the fact that everything is interconnected might result in wide-spread dependent failure behaviors among the interconnected systems, further impairing their reliability. Moreover, the transition into industry 4.0 often involves introducing many new technologies, which could bring another reliability problem, as in general, the potential failure modes/mechanisms are not well understood.

Along with the challenges also come opportunities. More and more data are becoming available as infrastructures get upgraded for industry 4.0. These data might contain useful reliability information. Meanwhile, more and more powerful computational tools/algorithms are emerging for extracting useful information from big data to support decision making [168]. In the next stage of my research, I intend to explore the possibility of applying these new big data analytic methods to better support understanding, modeling, and eventually improving the reliability of new industrial systems in industry 4.0. The overall objective of these researches is to develop a framework that allows integrating data of different nature to support online assessment and dynamic decision-making to improve the reliability of industry 4.0 applications. Three research projects are designed to achieve the overall research objective (detailed in the subsequent subsections).

11.1.1 Reliability modelling, analysis and prediction of cyber-physical systems based on stochastic hybrid systems

Cyber-physical systems are cornerstones for Industry 4.0 [94]. In cyber-physical systems, modern control systems and embedded software systems (cyber systems) are disposed of with an Internet address to be connected and addressed via IoT (the Internet of Things). Physical systems measure their current states, report it to the cyber systems, and react according to the control signals generated by the cyber systems, based on the physical system's current states. With the help of cyber-physical systems, key concepts in Industry 4.0 like smart factory can be realized.

An essential feature of cyber-physical systems is that they often exhibit mixtures of discrete and continuous behaviors [94]. The discrete behaviors mainly result from the control logic in the cyber systems, while the continuous behaviors are primarily a result of the physical systems' continuous dynamics. New models are needed to capture this feature to investigate the reliability of cyber-physical systems better. Our previous research discovered a stochastic hybrid system as a useful mathematical tool for describing systems' stochastic dynamics involving both discrete and continuous behaviors [44]. The most promising feature of the stochastic hybrid system is that a semi-analytical solution exists by transforming the system into differential equations, which could significantly improve the analysis's efficiency. In this project, we intend to extend our previous research and develop a stochastic hybrid system-based framework for modeling, analyzing, and finally improving cyber-physical systems' reliability. More specifically, this project focuses on:

- developing stochastic hybrid system models to describe the normal and failure behaviors of cyber-physical systems;
- developing analytical/semi-analytical approach for efficient reliability assessment;
- developing methods for updating system states and predicting future state and remaining useful life, based on online-collected data and information;
- developing optimization models to improve the reliability of cyber-physical systems.

11.1.2 Exploring unknown failures through knowledge graph: Using past lessons to prepare for new challenges

From a reliability perspective, Industry 4.0 will bring large amounts of unknown potential failures. As enabling technologies for Industry 4.0, systems need to be connected and communicate with one another, creating a huge complex system with a large number of inter-dependencies. New failure mode might appear at the system level as an ”emergence phenomena” of complex systems. Further, as Industry 4.0 often requires introducing many new

technologies, unknown failure modes and mechanisms might also appear since these new technologies are less matured than long-existing ones.

To ensure reliable and safe implementation of Industry 4.0, these potential unknown failures need to be identified and handled correctly in the design phase. A possible solution is to predict the unknown failures based on experience of similar systems. To do this, one needs (1) a structured approach to organize knowledge on failures of similar systems; (2) an efficient method to reason what failure could happen in the new systems based on the knowledge on the old systems.

A knowledge graph is a knowledge base that uses a graph-structured data model or topology to integrate data [90]. Knowledge graphs are often used to store interlinked descriptions of entities – objects, events, situations, or abstract concepts – with free-form semantics. Based on the knowledge graphs, logical inference can be performed, generating new knowledge based on the existing one. Hence, the knowledge graph is a promising tool to support exploring unknown failures of Industry 4.0. This project intends to develop a methodological framework supporting learning from experience and predicting potential failure patterns for newly developed systems operated in a new environment. More specifically, this project intends to:

- develop knowledge graph models to describe knowledge regarding failures in similar domains;
- develop prediction algorithm to predict the potential failure modes for the a newly designed based on the knowledge graph;
- use knowledge graph to investigate potential failure emergence phenomena in large-scale, interconnected cyber-physical systems.

11.1.3 Coordinated predictive maintenance planning for distributed cyber-physical systems

Industry 4.0 relies on large-scale, highly connected cyber-physical systems as enabling infrastructures. The high complexity and interdependence in these cyber-physical systems challenge not only their reliability design, but also their maintenance planning. The difficulties in maintenance planning mainly come from the following aspects:

1. Failure behaviors of the subsystems might depend on one another, making it difficult to derive an overall optimal maintenance plan.
2. Different stakeholders might operate different subsystems with different interests and goals for planning maintenance.
3. The scale of the system often brings in significant computational burden when performing maintenance planning.

This project intends to address these challenges by developing efficient algorithms for predictive maintenance planning of large-scale, interconnected cyber-physical systems. More specifically, this project intends to:

- develop remaining useful life prediction methods for distributed cyber-physical systems.
- develop an agent-based model to consider the behavior of each stakeholder in predictive maintenance planning.
- develop a reinforcement learning algorithm for optimal predictive maintenance planning to improve the computational efficiency.

11.1.4 Connection to my current research team

Currently, I work in an industrial chaire (Risk and Resilience of Complex Systems) financially supported by three companies, EDF, Orange and SNCF. The chaire focuses on researches and applications related to modeling, analysis and optimization of risk and resilience of complex systems. We have three permanent members in the team: Prof. Anne Barros, whose research interests include probabilistic and statistic methods for risk and resilience, degradation modeling, and predictive maintenance; Dr. Yiping Fang, whose research interests include decision-making under uncertainty, modeling and optimization of interdependent critical infrastructure; and myself). The three of us have shared research interests in a broad area of risk, reliability and resilience, while at the same time have different specific focus and skills that complement each other. By collaborating with them, I can benefit from their helps in stochastic modeling, optimization and predictive maintenance planning, which are great helps to implement my research projects describe before.

We have already benefited a lot from this inter-team collaboration. Together with Prof. Barros and Dr. Fang, I co-supervised four PhD students (on-going):

- Mr. Andrea Belle (Centralesupélec, France, Nov. 2019 - Present):
  - Thesis title: Resilience modelling and optimal protection planning for interconnected railway, electrical and telecommunication systems.
  - Co-supervised (50$\%$) with Prof. Anne Barros (50%).
- Mr. Youba Nait Belaid (Centralesupélec, France, Nov. 2019 - Present):
  - Thesis title: Resilience modelling of interdependent telecom and electrical networks (CIFRE EDF).
  - Co-supervised (33%) with Prof. Anne Barros (33%) and Dr. Yiping Fang (33%).
- Mr. Rui Li (Centralesupélec, France, Nov. 2020 - Present):
  - Thesis title: Resilience modelling and optimization for 5G infrastructures (CIFRE Orange).

  - Co-supervised (33%) with Prof. Anne Barros (33%) and Dr. Yiping Fang (33%).

- Mr. Khaled Sayad (Centralesupélec, France, Nov. 2020 - Present):
  - Thesis title: Joint optimization of maintenance activities considering interdependency in critical infrastructures (CIFRE Orange).
  - Co-supervised (33%) with Prof. Anne Barros (33%) and Dr. Yiping Fang (33%).

Some very interesting preliminary results have been achieved from these collaborations, although most of them are not published yet (this is why I did not include them in this thesis). These works can serve as solid foundations for the scientific projects described before. For example, working with Mr. Belle and Mr. Belaid, we had a few solid use cases on how to model a interconnected complex system, involving telecommunication between subsystems. This could serve as a prototype to continue investigating the modeling of cyber-physical system, as described in Sect. 11.1.1 and 11.1.3. The work of Mr. Li and Mr. Seyad could provide some prior knowledge when we are going to consider the coordination of maintenance in a distributed system as descried in Sect. 11.1.3.

If I could successfully obtain my HDR, it will also be a strong enhancement to the current team. Currently, we have only one HDR in the team (Prof. Barros). As Dr. Fang is also applying for HDR, in the ideal case, we are going to have three HDRs in the team soon. This means that we have opportunity to host more PhD projects, with different focuses but also synergies among the different projects. By doing so, as a team, we could explore more interesting challenges in the field of risk, reliability and resilience. For myself, I could continue benefiting from the collaborations within the team, which will significantly improve the stochastic modeling and optimization parts in my projects.

Chapter 12

CONCLUSIONS

This thesis summarizes the main research results obtained by me and the students I co-supervised, from Jan 2016 to present. My research activities in this period share a common goal, *i.e.,* developing new methods to improve the performance of risk and reliability analysis, under the practical constraints of limited historical failure data. To achieve this goal, new reliability models and analysis methods are developed, in which reliability is estimated based on failure behavior models, developed based on a deep understanding of the principles of failures, or dynamically assessed based on online collected data. More specifically, the research activities are conducted following five research axes:

- In research axis 1, a conceptual framework is developed for understanding common contributing factors to failures. Four main contributing factors to failure are identified, *i.e.,* design margin, aleatory uncertainty, degradation, and epistemic uncertainty. The difference between conscious and blind failure is highlighted, based on which the importance of considering epistemic uncertainty is explained. Finally, new risk and reliability metrics are defined that integrate all the contributing factors of the developed conceptual framework, especially the impact of epistemic uncertainty, for reliability quantification.

- In research axis 2, the focus is on modeling and analysis of degradation behavior, especially the degradation process involving both continuous performance degradation and discrete state changes. Dependency can also exist in the degradation process. A generic modeling framework is developed for such a dependent degradation process based on stochastic hybrid automaton. To improve the efficiency of the reliability assessment, a semi-analytical approach is developed for the degradation process that could be modeled by a special type of SHA, *i.e.,* the stochastic hybrid system. The developed modeling framework and analysis method is also extended to consider common-cause failure in the system-level.

- Research axis 3 focuses another important contributing factor to failure, the epistemic uncertainty. The main issues addressed in this axis is the practical assessment of epistemic uncertainty and its integration with

the risk assessment result. A maturity model is developed to describe the capability of epistemic uncertainty management, based on which practical assessment of epistemic uncertainty can be conducted through predefined scoring guidelines. A machine learning-based method is also developed for the assessment of epistemic uncertainty in practice. Finally, the integration of epistemic uncertainty is addressed by developing a new approach for multi-hazard risk aggregation based on Bayesian model averaging.

- In research axis $4,$ we consider systems with multi-state, recoverable behaviors. The focus of our research is to develop modeling framework for such systems that allows considering the costs during the recovery process. A Markov reward process-based model is developed for this and used to model the resilience of multi-state systems. To consider the time-dependent system behavior, the Markov reward process-based model is extended to a non-homogeneous semi-Markov reward-process based model. An efficient Monte Carlo simulation algorithm is also developed to improve the computational efficiency of the resilience analysis based on the developed model.

- Research axis 5 considers the integration of different, heterogeneous data sources for online reliability assessment and remaining useful life prediction. A sequential Bayesian updating algorithm is developed to integrate statistical failure data with condition-monitoring data, based on a hierarchical Bayesian model. The fusion of condition-monitoring and inspection data is also considered by developing a hidden Markov Gaussian mixture model and a data integration model based on Bayesian network. Expert judgment is also considered and integrated with the condition-monitoring data through an evidential hidden Markov model.

From a quantitative point of view, as of Jan 2022, I am the author/co-author of 32 papers in well-recognized international journals, 23 papers in international conferences, 2 book chapters (in Chinese) and 1 monograph (in Chinese). My research has gained wide community recognition. I have been nominated as editorial board member of International Journal of Data Analysis Techniques and Strategies since 2020, and technical committee member of a lot of important conferences including ESREL, ICSRS and ICRERM. I was invited as a lead guest editor of the international journal Applied Science on topic of dependent failure modeling. I have leaded/participated a number of research grants, funded by both industry and government funding agencies, with a total amount of ~ 220 K€.

From a student supervision point of view, I have co-supervised 4 PhD students who have successfully defended, and another 6 whose theses are still on-going. The total fraction of supervision is 350%. I have also supervised 4 master students. Through my supervision, my students achieved 16 papers in well-recognized journals and 6 in international conferences, including one winning best presentation award of the conference. I have been also actively involved in teaching with over 300 hETD since 2016. I am co-head of the international master program Risk, Resilience and Engineering Management (RREM) of Université Paris-Saclay.

All these activities impose a scientific rigor that I can further develop not only in my personal scientific production and in my teaching responsibilities, but also in the doctoral supervision and future works in order to establish our

research activity on solid qualitative bases. In the future, I plan to continue working on developing new models and approaches for understanding failure behavior and quantifying reliability, but on new systems and problems emerging in the context of industry 4.0. More specifically, the following projects are considered:

- reliability modeling, analysis and prediction of cyber-physical systems based on stochastic hybrid systems;
- exploring blind failures through knowledge graph: using past lessons to prepare for new challenges;
- coordinated predictive maintenance planning for distributed cyber-physical systems.

With the HDR, I will work more independently as a principle investigator, establish and lead a research team to better working on this topics.

Bibliography

[1] Tous-les-formulaires-hdr. https://www.universite-paris-saclay.fr/sites/default/files/2020-10/Tous-les-formulaires-HDR.docx. Accessed: 2020-12-30.

[2] Electricity price statistics. https://ec.europa.eu/eurostat/statistics-explained/index.php/Electricity_price_statistics. Accessed: 2019-04-22.

[3] Naive bayes, guassian distributions and practical applicationss. http://http://www.cs.cmu.edu/~tom/10601_sp09/lectures/NBayes2_2-2-2009-ann.pdf. Accessed: 2016-10-24.

[4] Politique de l'université paris-saclay en matière d'habilitation à diriger des recherches (hdr) et de dérogations à l'hdr. https://www.universite-paris-saclay.fr/sites/default/files/2020-10/Politique-HDR.pdf. Accessed: 2020-12-30.

[5] Quantitative risk assessment: Final report (prepared for nw innovation works, by acutech consulting group). http://http://kalamamfgfacilitysepa.com/wp-content/uploads/2016/09/FEIS-Appendix-G1-Quantitative-Risk-Assessment.pdf. Accessed: 2016-11-08.

[6] Regulatory guide 1.200: An approach for determining the technical adequacy of probabilistic risk assessment results for risk-informed activities. Technical report, US Nuclear Regulatory Commission, 2009.

[7] Earthquake preparedness and response for nuclear power plants. Technical report, International Atomic Energy Agency (IAEA), 2011.

[8] B. Allenby and J. Fink. Social and ecological resilience: toward inherently secure and resilient societies. *Science*, 24(3):347–364, 2000.

[9] T. Anagnos and A. S. Kiremidjian. A review of earthquake occurrence models for seismic hazard analysis, 1988.

[10] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. *IEEE Transactions on Signal Processing*, 50(2):174–188, 2002.

[11] T. Aven. On the meaning of a black swan in a risk context. *Safety science*, 57:44–51, 2013.

[12] T. Aven and B. Heide. Reliability and validity of risk analysis. *Reliability Engineering & System Safety*, 94(11): 1862–1868, 2009.

[13] T. Aven and R. Steen. The concept of ignorance in a risk assessment and risk management context. *Reliability Engineering & System Safety*, 95(11):1117–1122, 2010.

[14] T. Aven, P. Baraldi, R. Flage, and E. Zio. *Uncertainty in risk assessment: the representation and treatment of uncertainties by probabilistic and non-probabilistic methods*. John Wiley & Sons, 2013.

[15] H.-R. Bae, R. V. Grandhi, and R. A. Canfield. Epistemic uncertainty quantification techniques including evidence theory for large-scale structures. *Computers & Structures*, 82(13-14):1101–1112, 2004.

[16] V. Bagdonavičius, A. Bikelis, and V. Kazakevičius. Statistical analysis of linear degradation and failure time data with multiple failure modes. *Lifetime data analysis*, 10(1):65–81, 2004.

[17] J. W. Baker. An introduction to probabilistic seismic hazard analysis. *White paper, version*, 1:72, 2008.

[18] T. Bani-Mustafa, Z. Zeng, E. Zio, and D. Vasseur. A new framework for multi-hazards risk aggregation. *Safety science*, 121:283–302, 2020.

[19] M. Bao, Y. Ding, C. Singh, and C. Shao. A multi-state model for reliability assessment of integrated gas and power systems utilizing universal generating function techniques. *IEEE Transactions on Smart Grid*, 10(6): 6271–6283, 2019.

[20] P. Baraldi, F. Mangili, and E. Zio. A kalman filter-based ensemble approach with application to turbine creep prognostics. *IEEE Transactions on Reliability*, 61(4):966–977, 2012.

[21] R. E. Barlow and F. Proschan. *Mathematical theory of reliability*. SIAM, 1996.

[22] D. W. Benbow and H. W. Broome. *The Certified Reliability Engineer Handbook*. ASQ Quality Press, 2012.

[23] D. Bose, C. K. Chanda, and A. Chakrabarti. Vulnerability assessment of a power transmission network employing complex network theory in a resilience framework. *Microsystem Technologies*, pages 1–9, 2020.

[24] M. Bruneau, S. E. Chang, R. T. Eguchi, G. C. Lee, T. D. O'Rourke, A. M. Reinhorn, M. Shinozuka, K. Tierney, W. A. Wallace, and D. Von Winterfeldt. A framework to quantitatively assess and enhance the seismic resilience of communities. *Earthquake spectra*, 19(4):733–752, 2003.

[25] C. Carlson. *Effective FMEAs: Achieving safe, reliable, and economical products and processes using failure mode and effects analysis*, volume 1. John Wiley and Sons, 2012.

[26] G. P. Castaneda, J.-F. Aubry, and N. Brinzei. Stochastic hybrid automata model for dynamic reliability assessment. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 225(1): 28–41, 2011.

[27] J. H. Cha and M. Finkelstein. On new classes of extreme shock models and some generalizations. *Journal of Applied Probability*, 48(1):258–270, 2011.

[28] L. Chen. Safety of Nuclear Energy: Analysis of Events at Commercial Nuclear Power Plants. Master's thesis, ETH Zurich, Switzerland, 2018.

[29] L. Cloth and B. R. Haverkort. Five performability algorithms: A comparison. In *Markov Anniversary Meeting, Charleston, SC, USA*, pages 39–54, 2006.

[30] J. Collins, H. Busby, and G. Staab. *Mechanical design of machine elements and machines*. Wiley, 2009.

[31] J. A. Collins. *Failure of materials in mechanical design: analysis, prediction, prevention*. John Wiley & Sons, 1993.

[32] M. Compare, F. Martini, S. Mattafirri, F. Carlevaro, and E. Zio. Semi-markov model for the oxidation degradation mechanism in gas turbine nozzles. *IEEE Transactions on Reliability*, 65(2):574–581, 2016.

[33] M. J. Cushing, D. E. Mortin, T. J. Stadterman, and A. Malhotra. Comparison of electronics-reliability assessment approaches. *IEEE Transactions on reliability*, 42(4):542–546, 1993.

[34] A. Der Kiureghian and O. Ditlevsen. Aleatory or epistemic? does it matter? *Structural safety*, 31(2):105–112, 2009.

[35] J. Dezert, J.-M. Tacnet, M. Batton-Hubert, and F. Smarandache. Multi-criteria decision making based on dsmt-ahp. In *BELIEF 2010: Workshop on the Theory of Belief Functions*, pages 8–p. Belief Functions and Applications Society (BFAS), 2010.

[36] D. Draper. Assessment and propagation of model uncertainty. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1):45–70, 1995.

[37] E. L. Droguett and A. Mosleh. Bayesian methodology for model uncertainty using model performance data. *Risk Analysis: An International Journal*, 28(5):1457–1476, 2008.

[38] E. L. Droguett and A. Mosleh. Integrated treatment of model and parameter uncertainties through a bayesian approach. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 227 (1):41–54, 2013.

[39] S. Du, Z. Z. Zeng, Y.-p. Fang, and Q. Zhai. Resilience analysis of multistate systems based on markov reward processes. In *4th International Conference on System Reliability and Safety*, pages 12–17, 2019.

[40] J. Fan, S. Ghurye, and R. A. Levine. Multicomponent lifetime distributions in the presence of ageing. *Journal of applied probability*, 37(2):521–533, 2000.

[41] M. Fan. *Modeling, analysis and dynamic assessment of dependent failure processes*. PhD thesis, Beihang University, Beijing, China, 2019. In Chinese.

[42] M. Fan, Z. Zeng, E. Zio, and R. Kang. Modeling dependent competing failure processes with degradation-shock dependence. *Reliability Engineering & System Safety*, 165:422–430, 2017.

[43] M. Fan, Z. Zeng, E. Zio, R. Kang, and Y. Chen. A stochastic hybrid systems based framework for modeling dependent failure processes. *PloS one*, 12(2):e0172680, 2017.

[44] M. Fan, Z. Zeng, E. Zio, R. Kang, and Y. Chen. A stochastic hybrid systems model of common-cause failures of degrading components. *Reliability Engineering & System Safety*, 172:159–170, 2018.

[45] A. G. Frank, L. S. Dalenogare, and N. F. Ayala. Industry 4.0 technologies: Implementation patterns in manufacturing companies. *International Journal of Production Economics*, 210:15–26, 2019.

[46] F. Grabski. *Semi-Markov processes: applications in system reliability and maintenance*. Elsevier, 2014.

[47] F. Groen and A. Mosleh. Behavior of weighted likelihood and weighted posterior methods for treatment of uncertain data. In *Proc. ESREL*, volume 99, 1999.

[48] Y. Y. Haimes. On the definition of resilience in systems. *Risk Analysis: An International Journal*, 29(4): 498–501, 2009.

[49] P. Hall and J. Strutt. Probabilistic physics-of-failure models for component reliabilities using monte carlo simulation and weibull analysis: a parametric study. *Reliability Engineering & System Safety*, 80(3):233–242, 2003.

[50] M. S. Hamada, A. Wilson, C. S. Reese, and H. Martz. *Bayesian reliability*. Springer Science & Business Media, 2008.

[51] Y. Hao, X. Rong, H. Lu, Z. Xiong, and X. Dong. Quantification of margins and uncertainties for the risk of water inrush in a karst tunnel: representations of epistemic uncertainty with probability. *Arabian Journal for Science and Engineering*, 43(4):1627–1640, 2018.

[52] T. Hashimoto, J. R. Stedinger, and D. P. Loucks. Reliability, resiliency, and vulnerability criteria for water resource system performance evaluation. *Water resources research*, 18(1):14–20, 1982.

[53] W. He, N. Williard, M. Osterman, and M. Pecht. Prognostics of lithium-ion batteries based on dempster–shafer theory and the bayesian monte carlo method. *Journal of Power Sources*, 196(23):10314–10321, 2011.

[54] J. C. Helton and J. D. Johnson. Quantification of margins and uncertainties: Alternative representations of epistemic uncertainty. *Reliability Engineering & System Safety*, 96(9):1034–1052, 2011.

[55] J. P. Hespanha. Stochastic hybrid systems: Application to communication networks. In *International Workshop on Hybrid Systems: Computation and Control*, pages 387–401. Springer, 2004.

[56] J. P. Hespanha. A model for stochastic hybrid systems with application to communication networks. *Nonlinear Analysis: Theory, Methods & Applications*, 62(8):1353–1383, 2005.

[57] J. P. Hespanha. Modelling and analysis of stochastic hybrid systems. *IEE Proceedings-Control Theory and Applications*, 153(5):520–535, 2006.

[58] C. S. Holling. Resilience and stability of ecological systems. *Annual review of ecology and systematics*, 4(1): 1–23, 1973.

[59] S. Hosseini, K. Barker, and J. E. Ramirez-Marquez. A review of definitions and measures of system resilience. *Reliability Engineering & System Safety*, 145:47–61, 2016.

[60] Y. Hu, P. Baraldi, F. Di Maio, and E. Zio. A particle filtering and kernel smoothing-based approach for new design component prognostics. *Reliability Engineering & System Safety*, 134:19–31, 2015. doi: $10.1016/\mathrm{j.ress.2014.10.003}$.

[61] D. Huang, T. Chen, and M.-J. J. Wang. A fuzzy set approach for event tree analysis. *Fuzzy sets and systems*, 118(1):153–165, 2001.

[62] H.-Z. Huang and Z.-W. An. A discrete stress-strength interference model with stress dependent strength. *IEEE Transactions on Reliability*, 58(1):118–122, 2008.

[63] K. T. Huynh, A. Barros, C. Bérenguer, and I. T. Castro. A periodic inspection and replacement policy for systems subject to competing failure modes due to degradation and traumatic events. *Reliability Engineering & System Safety*, 96(4):497–508, 2011.

[64] E. IEC. International electrotechnical vocabulary (iev)—chapter 191—dependability and quality of service. *Web site http://www. electropedia. org*, 1990.

[65] T. Igusa, S. Buonopane, and B. Ellingwood. Bayesian analysis of uncertainty for structural engineering applications. *Structural Safety*, 24(2-4):165–186, 2002.

[66] J. Janssen and R. Manca. Numerical solution of non-homogeneous semi-markov processes in transient case. *Methodology and Computing in Applied Probability*, 3(3):271–293, 2001.

[67] K. Javed, R. Gouriveau, N. Zerhouni, and P. Nectoux. Enabling health monitoring approach based on vibration data for accurate prognostics. *IEEE Transactions on industrial electronics*, 62(1):647–656, 2014.

[68] D. Jenkinson. The elicitation of probabilities: A review of the statistical literature. Technical report, Citeseer, 2005.

[69] L. Jiang, Q. Feng, and D. W. Coit. Reliability and maintenance modeling for dependent competing failure processes with shifting failure thresholds. *IEEE Transactions on Reliability*, 61(4):932–948, 2012.

[70] L. Jiang, Q. Feng, and D. W. Coit. Modeling zoned shock effects on stochastic degradation in dependent failure processes. *Iie Transactions*, 47(5):460–470, 2015.

[71] S. Jiang, W. Zhang, J. He, and Z. Wang. Comparative study between crack closure model and willenborg model for fatigue prediction under overload effects. *Chinese Journal of Aeronautics*, 2016.

[72] T. Jiang, Y. Liu, and Y.-X. Zheng. Optimal loading strategy for multi-state systems: Cumulative performance perspective. *Applied Mathematical Modelling*, 74:199–216, 2019.

[73] L. Jiao, Q. Pan, Y. Liang, X. Feng, and F. Yang. Combining sources of evidence with reliability and importance for decision making. *Central European Journal of Operations Research*, 24(1):87–106, 2016.

[74] G. Jin, D. Matthews, Y. Fan, and Q. Liu. Physics of failure-based degradation modeling and lifetime prediction of the momentum wheel in a dynamic covariate environment. *Engineering Failure Analysis*, 28:222–240, 2013.

[75] C. Kai-Yuan, W. Chuan-Yuan, and Z. Ming-Lian. Fuzzy variables as a basis for a theory of fuzzy reliability in the possibility context. *Fuzzy sets and systems*, 42(2):145–172, 1991.

[76] M. Kalantarnia, F. Khan, and K. Hawboldt. Dynamic risk assessment using failure assessment and bayesian theory. *Journal of Loss Prevention in the Process Industries*, 22(5):600–606, 2009. doi: $10.1016/\mathrm{j.jlp.2009.04.006}$.

[77] S. Kaplan and B. J. Garrick. On the quantitative definition of risk. *Risk analysis*, 1(1):11–27, 1981.

[78] R. W. Katz. Stochastic modeling of hurricane damage. *Journal of Applied Meteorology*, 41(7):754–762, 2002.

[79] R. Kazemi and A. Mosleh. Improving default risk prediction using bayesian model uncertainty techniques. *Risk Analysis: An International Journal*, 32(11):1888–1900, 2012.

[80] D. Kelly and C. Smith. *Bayesian inference for probabilistic risk assessment: a practitioner's guidebook*. Springer Science & Business Media, 2011.

[81] D. L. Kelly and C. L. Smith. Bayesian inference in probabilistic risk assessment—the current state of the art. *Reliability Engineering & System Safety*, 94(2):628–643, 2009.

[82] H. Kim, S.-H. Lee, J.-S. Park, H. Kim, Y.-S. Chang, and G. Heo. Reliability data update using condition monitoring and prognostics in probabilistic safety assessment. *Nuclear Engineering and Technology*, 47(2): 204–211, 2015. doi: $10.1016/\mathrm{j.net}.2014.12.008$.

[83] F. Lees. *Lees' Loss prevention in the process industries: Hazard identification, assessment and control*. Butterworth-Heinemann, 2012.

[84] Y. Lei, N. Li, L. Guo, N. Li, T. Yan, and J. Lin. Machinery health prognostics: A systematic review from data acquisition to rul prediction. *Mechanical systems and signal processing*, 104:799–834, 2018.

[85] W. Li and H. Pham. Reliability modeling of multi-state degraded systems with multi-competing failures and random shocks. *IEEE transactions on reliability*, 54(2):297–303, 2005.

[86] X.-Y. Li, W.-B. Chen, and R. Kang. Performance margin-based reliability analysis for aircraft lock mechanism considering multi-source uncertainties and wear. *Reliability Engineering & System Safety*, 205:107234, 2021.

[87] Y. Li and E. Zio. Uncertainty analysis of the adequacy assessment model of a distributed generation system. *Renewable Energy*, 41:235–244, 2012.

[88] Y.-F. Li and E. Zio. A multi-state model for the reliability assessment of a distributed generation system via universal generating function. *Reliability Engineering & System Safety*, 106:28–36, 2012.

[89] X. Liao. Research on the wear mechanism and life modeling method of aero-hydraulic spool valve. *Degree of Master, Beihang University, China*, 2014.

[90] Y. Lin, Z. Liu, M. Sun, Y. Liu, and X. Zhu. Learning entity and relation embeddings for knowledge graph completion. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 29, 2015.

[91] B. Liu. *Uncertainty Theory: A Branch of Mathematics for Modeling Human Uncertainty*. Springer-Verlag, Berlin, 2010.

[92] Y. Liu and C.-J. Chen. Dynamic reliability assessment for nonrepairable multistate systems by aggregating multilevel imperfect inspection data. *IEEE Transactions on Reliability*, 66(2):281–297, 2017.

[93] Y. Liu, P. Lin, Y.-F. Li, and H.-Z. Huang. Bayesian reliability and performance assessment for multi-state systems. *IEEE Transactions on Reliability*, 64(1):394–409, 2014.

[94] Y. Lu. Cyber physical system (cps)-based industry 4.0: A survey. *Journal of Industrial Integration and Management*, 2(03):1750014, 2017.

[95] G. M. Safety significance evaluation of kewaunee power station turbine building internal floods. dominion resources inc. 2005.

[96] G. Masala, G. Cannas, and M. Micocci. Survival probabilities for hiv infected patients through semi-markov processes. *Biometrical Letters*, 51(1):13–36, 2014.

[97] J. W. McPherson. *Reliability physics and engineering: time-to-failure modeling*. Springer, 2018.

[98] W. Q. Meeker, L. A. Escobar, and F. G. Pascual. *Statistical methods for reliability data*. John Wiley & Sons, 2021.

[99] A. Meel and W. D. Seider. Plant-specific dynamic failure assessment using bayesian theory. *Chemical Engineering Science*, 61(21):7036–7056, 2006. doi: 10.1016/j.ces.2006.07.007.

[100] A. Meel and W. D. Seider. Real-time risk analysis of safety systems. *Computers & Chemical Engineering*, 32 (4-5):827–840, 2008. doi: 10.1016/j.compchemeng.2007.03.006.

[101] Z. Mohaghegh and M. Modarres. A probabilistic physics-of-failure approach to common cause failures in reliability assessment of structures and components. *Transactions of the American Nuclear Society*, 105: 635–637, 2011.

[102] A. Mosleh and G. Apostolakis. The assessment of probability distributions from expert opinions with an application to seismic fragility curves. *Risk analysis*, 6(4):447–461, 1986.

[103] R. Munikoti and P. Dhar. Highly accelerated life testing (halt) for multilayer ceramic capacitor qualification. *IEEE Transactions on Components, Hybrids, and Manufacturing Technology*, 11(4):342–345, 1988.

[104] H. Nabli and B. Sericola. Performability analysis: a new algorithm. *IEEE Transactions on Computers*, 45(4): 491–494, 1996.

[105] H.-P. Nguyen, J. Liu, and E. Zio. Dynamic-weighted ensemble for fatigue crack degradation state prediction. *Engineering Fracture Mechanics*, 194:212–223, 2018.

[106] J. S. Nielsen and J. D. Sørensen. Bayesian estimation of remaining useful life for wind turbine blades. *Energies*, 10(5):664, 2017.

[107] NUREG1855. Guidance on the treatment of uncertainties associated with pras in risk informed decision making. Technical report, Nuclear Researc Commission, 2013.

[108] M. Panteli and P. Mancarella. Modeling and evaluating the resilience of critical electrical power infrastructure to extreme weather events. *IEEE Systems Journal*, 11(3):1733–1742, 2017.

[109] M. Pecht and A. Dasgupta. Physics-of-failure: an approach to reliable product development. In *Integrated Reliability Workshop*, Integrated Reliability Workshop, pages pp. 1–4, Lake Tahoe, US, 1995. IEEE. Integrated Reliability Workshop, 1995. Final Report., International.

[110] N. Pedroni, E. Zio, A. Pasanisi, and M. Couplet. A critical discussion and practical recommendations on some issues relevant to the nonprobabilistic treatment of uncertainty in engineering risk assessment. *Risk Analysis*, 37(7):1315–1340, 2017.

[111] H. Peng, Q. Feng, and D. W. Coit. Reliability and maintenance modeling for systems subject to multiple dependent competing failure processes. *IIE transactions*, 43(1):12–22, 2010.

[112] H. Pishro-Nik. Introduction to probability, statistics, and random processes. 2016.

[113] Z. Qiu, R. Huang, X. Wang, and W. Qi. Structural reliability analysis and reliability-based design optimization: Recent advances. *Science China Physics, Mechanics and Astronomy*, 56(9):1611–1618, 2013.

[114] S. Raadnui and S. Kleesuwan. Low-cost condition monitoring sensor for used oil analysis. *Wear*, 259(7-12): 1502–1506, 2005.

[115] L. R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

[116] A. Rae, R. Alexander, and J. McDermid. Fixing the cracks in the crystal ball: A maturity model for quantitative risk assessment. *Reliability Engineering & System Safety*, 125:67–81, 2014. doi: $10.1016/\mathrm{j.ress.2013.09.008}$.

[117] K. Rafiee, Q. Feng, and D. W. Coit. Reliability modeling for dependent competing failure processes with changing degradation rate. *IIE transactions*, 46(5):483–496, 2014.

[118] E. Ramasso and T. Denoeux. Making use of partial knowledge about hidden states in hmms: an approach based on belief functions. *IEEE Transactions on Fuzzy Systems*, 22(2):395–405, 2013.

[119] M. E. Riley and R. V. Grandhi. Quantification of model-form and predictive uncertainty for multi-physics simulation. *Computers & structures*, 89(11-12):1206–1213, 2011.

[120] R. Rocchetta, E. Zio, and E. Patelli. A power-flow emulator approach for resilience assessment of repairable power grids subject to weather-induced failures and data deficiency. *Applied energy*, 210:339–350, 2018.

[121] S. M. Ross. *Introduction to probability models*. Academic press, 2014.

[122] T. L. Saaty. Decision making with the analytic hierarchy process. *International journal of services sciences*, 1 (1):83–98, 2008.

[123] B. Saha and K. Goebel. Battery data set. *NASA AMES prognostics data repository*, 2007.

[124] A. Sasaki and T. Yamamoto. A review of studies of hydraulic lock. *Lubrication engineering*, 49(8):585–593, 1993.

[125] G. Shafer. *A mathematical theory of evidence*. Princeton university press, 1976.

[126] Y. Sheffi, D. Closs, J. Davidson, D. French, B. Gordon, R. Martichenko, J. Mentzer, C. Norek, N. Seiersen, and T. Stank. Supply chain resilience how can you transcend vulnerability in your supply chain to gain competitive advantage. *The Official Magazine of The Logistics Institute*, 12(1):12–17, 2006.

[127] Y. Sheffi et al. The resilient enterprise: overcoming vulnerability for competitive advantage. *MIT Press Books*, 1, 2005.

[128] X.-S. Si, W. Wang, C.-H. Hu, and D.-H. Zhou. Remaining useful life estimation–a review on the statistical data driven approaches. *European journal of operational research*, 213(1):1–14, 2011.

[129] P. Smets and R. Kennes. The transferable belief model. *Artificial intelligence*, 66(2):191–234, 1994.

[130] H. C. Tijms and R. Veldman. A fast algorithm for the transient reward distribution in continuous-time markov chains. *Operations Research Letters*, 26(4):155–158, 2000.

[131] D. A. Tobon-Mejia, K. Medjaher, N. Zerhouni, and G. Tripot. A data-driven failure prognostics method based on mixture of gaussians hidden markov models. *IEEE Transactions on reliability*, 61(2):491–503, 2012.

[132] P. Todorovic and E. Zelenhasic. A stochastic model for flood analysis. *Water Resources Research*, 6(6): 1641–1648, 1970.

[133] C. W. Tsai, N.-K. Wu, and C.-H. Huang. A multiple-state discrete-time markov chain model for estimating suspended sediment concentrations in open channel flow. *Applied Mathematical Modelling*, 40(23-24):10002–10019, 2016.

[134] E. D. Vugrin, D. E. Warren, M. A. Ehlen, and R. C. Camphouse. A framework for assessing the resilience of infrastructure and economic systems. In *Sustainable and resilient critical infrastructure systems*, pages 77–116. Springer, 2010.

[135] J. Wang, W. Zuo, L. Rhode-Barbarigos, X. Lu, J. Wang, and Y. Lin. Literature review on modeling and simulation of energy infrastructures from a resilience perspective. *Reliability Engineering & System Safety*, 2018.

[136] Y. Wang and H. Pham. A multi-objective optimization of imperfect preventive maintenance policy for dependent competing risk systems with hidden failure. *IEEE Transactions on Reliability*, 60(4):770–781, 2011.

[137] P. Withey. Fatigue failure of the de havilland comet i. *Engineering failure analysis*, 4(2):147–154, 1997.

[138] Z. X, Y. A, and T. T. Quantitative common cause failure modeling for auxiliary feedwater system involving the seismic-induced degradation of flood barriers. 51:332–342, 2014.

[139] T. Xiahou, Z. Zeng, and Y. Liu. Remaining useful life prediction by fusing expert knowledge and condition monitoring information. *IEEE Transactions on Industrial Informatics*, 17(4):2653–2663, 2020.

[140] C. Xie, G. Li, and F. Wei. An integrated qmu approach to structural reliability assessment based on evidence theory and kriging model with adaptive sampling. *Reliability Engineering & System Safety*, 171:112–122, 2018.

[141] J. Xing, Z. Zeng, and E. Zio. A framework for dynamic risk assessment with condition monitoring data and inspection data. *Reliability Engineering & System Safety*, 191:106552, 2019.

[142] G. Yang. *Life cycle reliability engineering*. John Wiley and Sons,, Hoboken, N.J. :, 2007.

[143] X. Yang, Y. Liu, Y. Zhang, and Z. Yue. Hybrid reliability analysis with both random and probability-box variables. *Acta Mechanica*, 226(5):1341–1357, 2015.

[144] Y.-J. Yang, W. Peng, D. Meng, S.-P. Zhu, and H.-Z. Huang. Reliability analysis of direct drive electrohydraulic servo valves based on a wear degradation process and individual differences. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 228(6):621–630, 2014.

[145] Z. S. Ye, L. C. Tang, and H. Y. Xu. A distribution-based systems reliability model under extreme shocks and natural degradation. *IEEE Transactions on Reliability*, 60(1):246–256, 2011.

[146] B. D. Youn, C. Hu, and P. Wang. Resilience-driven system design of complex engineered systems. *Journal of Mechanical Design*, 133(10):101011, 2011.

[147] O. Zadakbar, S. Imtiaz, and F. Khan. Dynamic risk assessment and fault detection using principal component analysis. *Industrial & Engineering Chemistry Research*, 52(2):809–816, 2012.

[148] O. Zadakbar, S. Imtiaz, and F. Khan. Dynamic risk assessment and fault detection using a multivariate technique. *Process Safety Progress*, 32:365–375, 2013.

[149] O. Zadakbar, F. Khan, and S. Imtiaz. Dynamic risk assessment of a nonlinear non-gaussian system using a particle filter and detailed consequence analysis. *The Canadian Journal of Chemical Engineering*, 93(7): 1201–1211, 2015. doi: 10.1002/cjce.22212.

[150] Z. Zeng and E. Zio. A classification-based framework for trustworthiness assessment of quantitative risk analysis. *Safety Science*, 99:215–226, 2017.

[151] Z. Zeng and E. Zio. An integrated modeling framework for quantitative business continuity assessment. *Process Safety and Environmental Protection*, 106:76–88, 2017.

[152] Z. Zeng and E. Zio. Dynamic risk assessment based on statistical failure data and condition-monitoring degradation data. *IEEE Transactions on Reliability*, 67(2):609–622, 2018.

[153] Z. Zeng, M. Wen, and R. Kang. Belief reliability: A new metrics for products' reliability. *Fuzzy Optimization and Decision Making*, 12(1):15–27, 2013.

[154] Z. Zeng, R. Kang, and Y. Chen. A physics-of-failure-based approach for failure behavior modeling: With a focus on failure collaborations, 2014.

[155] Z. Zeng, R. Kang, M. Wen, and Y. Chen. Measuring reliability during product development considering aleatory and epistemic uncertainty, 2015.

[156] Z. Zeng, R. Kang, and Y. Chen. Using pof models to predict system reliability considering failure collaboration. *Chinese Journal of Aeronautics*, 29(5):1294–1301, 2016.

[157] Z. Zeng, Y. Chen, E. Zio, and R. Kang. A compositional method to model dependent failure behavior based on pof models. *Chinese Journal of Aeronautics*, 2017.

[158] Z. Zeng, R. Kang, M. Wen, and E. Zio. A model-based reliability metric considering aleatory and epistemic uncertainty. *IEEE Access*, 5:15505–15515, 2017.

[159] Z. Zeng, R. Kang, M. Wen, and E. Zio. Uncertainty theory as a basis for belief reliability. *Information Sciences*, 429:26–36, 2018.

[160] Z. Zeng, T. Bani-Mustafa, R. Flage, and E. Zio. An integrated risk index accounting for epistemic uncertainty in probability risk assessment. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, page 1748006X20968954, 2020.

[161] Z. Zeng, S. Du, and Y. Ding. Resilience analysis of multi-state systems with time-dependent behaviors. *Applied Mathematical Modelling*, 90:889–911, 2021.

[162] Z. Zeng, Y.-P. Fang, Q. Zhai, and S. Du. A markov reward process-based framework for resilience analysis of multistate energy systems under the threat of extreme events. *Reliability Engineering & System Safety*, 209: 107443, 2021.

[163] G. Zhai, Y. Zhou, and X. Ye. A tolerance design method for electronic circuits based on performance degradation. *Quality and Reliability Engineering International*, 31(4):635–643, 2015.

[164] Q. Zhang, F. Zheng, Q. Chen, Z. Kapelan, K. Diao, K. Zhang, and Y. Huang. Improving the resilience of postdisaster water distribution systems using dynamic optimization framework. *Journal of Water Resources Planning and Management*, 146(2):04019075, 2020.

[165] Y.-G. Zhao and T. Ono. Moment methods for structural reliability. *Structural safety*, 23(1):47–75, 2001.

[166] C. Zheng, P. Ge, and Y. Li. Contaminant lock force and filter cake forming mechanism of hydraulic spool valves. *Lubrication Engineering*, pages 14–19, 2014.

[167] Z. Zhou. Research on mechanism of contaminant lock for hydraulic valve. *Chinese Hydraulics & Pneumatics*, 1(12):15–17, 1994.

[168] E. Zio. The future of risk assessment. *Reliability Engineering & System Safety*, 177:176–190, 2018.

[169] E. Zio and G. Apostolakis. Two methods for the structured assessment of model uncertainty by experts in performance assessments of radioactive waste repositories. *Reliability Engineering and System Safety*, 54 (2):225–241, 1996.

[170] E. Zio and G. Apostolakis. Two methods for the structured assessment of model uncertainty by experts in performance assessments of radioactive waste repositories. *Reliability Engineering & System Safety*, 54(2-3): 225–241, 1996.

université PARIS-SACLAY
ÉCOLE DOCTORALE
INTERFACES
Approches Interdisciplinaires : Fondements, Applications et Innovation

Titre: Rendre l'ingénierie de la fiabilité intelligente : lorsque les principes de défaillance rencontrent le Big Data industriel

Mots clés: Fiabilité, risque, résilience, incertitude, analyse de données

Résumé: Cette thèse résume mes principales activités de recherche de janvier 2016 à aujourd'hui pour soutenir ma candidature à l'Habilitation à Diriger des Recherches (HDR) à l'Université Paris-Saclay. L'objectif général de mes activités de recherche est d'améliorer les performances de l'analyse des risques et de la fiabilité, sous les contraintes pratiques de données de défaillance historiques limitées. Pour atteindre cet objectif, nous avons travaillé sur deux directions, *i.e.,* la modélisation du comportement de défaillance basée sur la physique et l'évaluation de la fiabilité dynamique à travers des données collectées en ligne. Plus précisément, cinq axes de recherche sont considérés : les cadres conceptuels pour les causes de défaillance, la modélisation du comportement de défaillance dépendante, la quantification de l'incertitude épistémique, la modélisation de la résilience d'un système multi-états et l'intégration de données multi-sources pour l'évaluation de la fiabilité. La thèse comprend deux parties. Dans la partie I, des résumés synthétiques de mes activités de recherche, d'enseignement et d'encadrement d'étudiants sont présentés pour soutenir l'application du RDH. Dans la partie II, les résultats de recherche représentatifs de chacun des cinq axes sont brièvement présentés.

Title: Making Reliability Engineering Smart: When Principles of Failure Meet with Industrial Big Data

Keywords: Reliability, risk, resilience, uncertainty, data analytics

Abstract: This thesis summarizes my main research activities from Jan 2016 to present to support my application to Habilitation à Diriger des Recherches (HDR) at Université Paris-Saclay. The general objective of my researc activities is to improve the performance of risk and reliability analysis, under the practical constraints of limited historical failure data. To achieve this goal, we worked on two directions, *i.e.,* physics-based failure behavior modeling and dynamic reliability assessment through online collected data. More specifically, five research axes are considered: conceptual frameworks for failure causes, dependent failure behavior modeling, quantification of epistemic uncertainty, resilience modeling of multi-state system, and multi-source data integration fro reliability assessment. The thesis comprises of two parts. In Part I, synthetic summaries of my research, teaching and student supervision activities are presented to support the application of HDR. In Part II, the representative research results from each of the five axes are briefly introduced.

Université Paris-Saclay
Laboratoire Genie Industriel, Centralesupélec
3 Rue Joliot Curie, 91190 Gif-sur-Yvette, France