Multi-Hazards Risk Aggregation Considering the Trustworthiness of the Risk Assessment

Tasneem Bani-Mustafa

NNT : 2019SACLC096

Multi-Hazards Risk Aggregation Considering the Trustworthiness of the Risk Assessment

Thèse de doctorat de l'Université Paris-Saclay
préparée à CentraleSupélec

École doctorale n° 573, Interfaces : approches interdisciplinaires / fondements, applications et innovation
Spécialité de doctoratt: Ingénierie des Systèmes Complexes

Thèse présentée et soutenue à Gif-Sur-Yvette, le 02/Dec/2019, par

Tasneem BANI-MUSTAFA

Composition du Jury :

Terje AVEN Professor, University of Stavanger (SEROS)	Rapporteur
Maria Francesca Milazzo Professor, Universita' degli Studi di Messina (Dipartimento di Ingegneria)	Rapporteur
Vincent Mousseau Professor, CentraleSupélec (laboratoire MICS)	Président du jury
Wassila Ouerdane Assisstant Professor, CentraleSupélec (laboratoire MICS)	Examinateur
François BEAUDOUIN Ingenieur chercheur, EDF (EDF R&D, PRISME)	Examinateur
Enrico Zio Professor, Politecnico di Milano (Energy Department)	Directeur de thèse
Zhiguo Zeng Assistant Professor, CentraleSupélec (Laboratoire LGI)	Encadrant
Dominique Vasseur Ingenieur chercheur, EDF (EDF R&D, PERICLES)	Encadrante

Thèse de doctorat

Titre : L'agrégation De Risques Multiples Pour Les Centrales Nucléaires Dans Le Contexte De L'aide A La Décision

Mots clés : Analyse de risque, Aide à la décision, Agrégation des risques multiples, Centrale nucléaire, Niveau de réalisme et confiance dans l'analyse de risque, Connaissance.

Résumé : Cette thèse de doctorat aborde le problème de l'agrégation de risques multiple (MHRA), qui vise à agréger les risques estimés pour différents contributeurs.
La pratique actuelle de la MHRA est basée sur une sommation arithmétique simple des estimations de risques. Cependant, ces estimations sont obtenues à partir de modèles EPS (Estimation Probabiliste de risque) qui présentent des degrés de réalisme différents liés à différents niveaux de connaissances. En ne prenant pas en compte ces différences, le processus MHRA pourrait conduire à des résultats trompeurs pour la prise de décision (DM). Dans cette thèse, un cadre structuré est proposé afin d'évaluer le niveau de réalisme et de confiance dans les évaluations de risques et de l'intégrer dans le processus de MHRA.
Ces travaux ont permis :
(i) Une identification des facteurs contribuant à la fiabilité de l'évaluation des risques. Leurs criticités sont analysées afin de comprendre leur influence sur l'estimation des risques;
(ii) Un cadre hiérarchique intégré est développé pour évaluer la confiance et le réalisme de l'estimation de risque, sur la base des facteurs et des attributs identifiés en (i);
(iii) Une méthode basée sur un modèle réduit est proposée pour évaluer efficacement la fiabilité de l'évaluation des risques dans la pratique. Grâce à cette méthode, le nombre d'éléments pris en compte dans l'évaluation initiale des risques peut être limité.
(iv) Une technique qui combine la théorie de Dempster-Shafer et le processus de hiérarchie analytique (DST-AHP) est appliquée au modèle développé. Cette technique permet d'évaluer le niveau de réalisme et confiance -dans l'analyse de risque- en utilisant une moyenne pondérée des attributs: la méthode AHP est utilisée pour calculer le poids des attributs et la méthode DST est utilisée pour tenir compte de l'incertitude subjective dans le jugement des experts dans l'évaluation des poids;
(v) Une technique de MHRA est développée sur la base d'un modèle de moyenne bayésienne afin de surmonter les limites de la pratique actuelle de MHRA qui néglige le réalisme et confiance dans l'évaluation de chaque contributoire de risque;
(vi) Le modèle développé est appliqué sur des cas réels de l'industrie des centrales nucléaires.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France

Title : Multi-Hazards Risk Aggregation considering the trustworthiness of the assessment

Keywords : Risk assessment, Multi-hazards risk aggregation (MHRA), Nuclear power plants, Trustworthiness in risk analysis, Background knowledge, Risk-informed decision-making

Abstract: This PhD thesis addresses the problem of Multi-Hazards Risk Aggregation (MHRA), which aims at aggregating the risk estimates from Probabilistic Risk Assessment (PRA) models for the different contributors. The current practice of MHRA is based on a simple arithmetic summation of the risk estimates. However, the risk estimates are obtained from PRA models that have different degrees of trustworthiness, because of the different background knowledge they are based on. Ignoring this difference in MHRA could lead to misleading results for Decision-Making (DM). In this thesis, a structured framework is proposed to assess the level of trustworthiness, which risk assessment results are based on and to integrate it in the process of MHRA.

The original scientific contributions are:

(i) Factors contributing to the trustworthiness of risk assessment outcomes are identified and their criticalities are analyzed under different frameworks, to understand their influence on the risk results;

(ii) An integrated hierarchical framework is developed for assessing the trustworthiness of risk analysis, based on the identified factors and related attributes;

(iii) A reduced order model-based method is proposed to efficiently evaluate the trustworthiness of risk assessment in practice. Through the reduced-order model, the proposed method can limit the number of elements considered in the original risk assessment;

(iv) A technique that combines Dempster Shafer Theory and the Analytic Hierarchy Process (namely, DST-AHP) is applied to the developed framework to assess the trustworthiness by a weighted average of the attributes in the framework: the AHP method is used to derive the weights of the attributes and the DST is used to account for the subjective uncertainty in the experts' judgments for the evaluation of the weights;

(v) A MHRA technique is developed based on Bayesian model averaging, to overcome the limitations of the current practice of risk aggregation that neglects the trustworthiness of the risk assessment of individual hazard groups;

(vi) The developed framework is applied to real case studies from the Nuclear Power Plants (NPP) industry.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France

“개천에서 용 난다”

“A dragon rises up from a small stream”

Korean proverb

“我们最大的荣耀不是永不堕落，而是每次跌倒时都会崛起”

“Our greatest glory is not in never falling, but in rising every time we fall”

Confucius

"وتحسب أنك جرم صغير وفيك انطوى العالم الاكبر"

“You think of yourself as a small orb, but, in fact, within you lies a great universe”

Ali Ben Abi-Taleb

‘Venez jusqu’au bord’, ‘Nous ne pouvons pas, nous avons peur.’
‘Venez jusqu’au bord.’, ‘Nous ne pouvons pas, nous allons tomber.’
Venez jusqu’au bord.’,
Et ils y sont allés
Et il les a poussés.
Et ils se sont envolés.

‘Come to the edge’, he said. "We can't, we're afraid!" they responded.
‘Come to the edge’, he said. ‘We can't, We will fall!’, they responded.
‘Come to the edge’, he said.
And so they came.
And he pushed them.
And they flew.

Guillaume Apollinaire

Acknowledgments

I would like to express my gratitude to all the people who contributed in one way or another to this thesis. First, I would like to thank the person, who challenged me the most, made this Ph.D. into such an interesting and beautiful adventure, my dear supervisor Zhiguo Zeng. Huge thanks go also to my supervisor at EDF, Dr. Dominique Vasseur. Thank you for believing in me from the very beginning, for making a huge effort throughout this Ph.D., and for being the kind human who you are. My deepest gratefulness also goes to the person who has "imprinted" on me, both professionally and personally, saw the best in me and was my northern star when I was lost in my darkest nights, my director and supervisor Prof. Enrico Zio. Thank you for being patient, visionary, for believing in me and setting the best goals for me. The quality of work would never have been the same if it was not for your guidance.

I had also, during this Ph.D., the honor of working with Prof. Roger Flage, who taught me a lot about the Scandinavian discipline and morals, with whom the work was very efficient and fruitful.

I'm grateful to my colleagues at my team, Dr. Yiping Fang, Dr. William Fauriat, with whom I had the most interesting discussions on all aspects. My friends Hoang-Phuong Nguyen, Daogui Tang, all the former colleagues at our team and my dearest comrade, Islam Abdin who made this journey a meaningful one. To my Jinduo Xing and Honping Wang who were more than sisters to me, the ones who taught me a new meaning of friendship. Thank you all for making these beautiful and unforgettable memories. You brought joy to my life.

I'm also thankful for all my current and former colleagues at LGI, I would like to thank Abood Mourad, Bruna Cavazza, Fabio Antonialli, Haythem Selmi, Hichem Benbitour, Gustavo Santamaria, Selmen Boubaker, Ouail Al Maghraoui, Reza Vosooghi, and for any person who had been a part of this lab and is not mentioned here. I am grateful to this small laboratory that brought together, people from all over the world, on one man's heart.

Many thanks also go to Carole Stoll, Corinne Olivier, Delphine Martin, Matthieu Tournadre, the head of this laboratory Prof. Bernard Yannou, and not to forget all the professors and researchers of this laboratory for immersing me with love and for being an affectionate and real family throughout these years.

To my friends who shared my days here in France, sweet and bitter ones, to whom my life would have never been at this ease without their support. To the love of my life Ghassen, who shared with me the most meaningful journey and made it a delightful unforgettable one. To my family in Jordan who have always been there for me despite the bitter of distance: my father and mother Adeeb and Faddah Bani-Mustafa, my siblings: Modafer, Thekra, Baraa, Abu-Baker, Ahmad, my sister and brothers in law: Afaf, Hani and Ashraf. My nephews: Hashem, Bana, Jad and Eyas. To those who swallowed their sorrow to make me happy and who believed in me from the very beginning. To my ever-after friends Hadeel Otoum, Hassanah Quqazeh, and Yazan Alatrash, those who have never changed despite the distance. Thank you all for being a part of my life and for supporting me.

To that person who might not remember this old pupil, yet, he believed in me a long time ago. I would not be more sincere if I said that you have participated a lot in what is written in this thesis. Thank you Dr. Samer Mautlq Ayasrah.

To that person who has changed my life, pushed me somehow to be standing where I am today, believed in me, taught me how to fall in love in what I am doing and to appreciate the journey more than the destination. To Prof. Mahmoud Elgohary: thank you, my teacher.

To “every single person” who passed through my life, you have participated in a way or another to who I am today. The achievement of this thesis is, all and entirely, yours. Thank you all for your unlimited and unconditional love and support. A new adventure will now begin and you will always be a part of me as you have always been.

Until a better achievement of yours,

Yours sincerely,

Tasneem Bani-Mustafa

France, 2019

Dedicated to all those who are fighting to make their dreams come true…..

To the old me, who had never given up….

To the future me, to whom I believe will be a better person than today's …..

Multi-Hazards Risk Aggregation considering the trustworthiness of the assessment

Abstract

This PhD thesis addresses the problem of Multi-Hazards Risk Aggregation (MHRA), which aims at aggregating the risk estimates from Probabilistic Risk Assessment (PRA) models for the different contributors. The current practice of MHRA is based on a simple arithmetic summation of the risk estimates. However, the risk estimates are obtained from PRA models that have different degrees of trustworthiness, because of the different background knowledge they are based on. Ignoring this difference in MHRA could lead to misleading results for Decision-Making (DM). In this thesis, a structured framework is proposed to assess the level of trustworthiness, which risk assessment results are based on and to integrate it in the process of MHRA.

The original scientific contributions are:

(i) Factors contributing to the trustworthiness of risk assessment outcomes are identified and their criticalities are analyzed under different frameworks, to understand their influence on the risk results;

(ii) An integrated hierarchical framework is developed for assessing the trustworthiness of risk analysis, based on the identified factors and related attributes;

(iii) A reduced order model-based method is proposed to efficiently evaluate the trustworthiness of risk assessment in practice. Through the reduced-order model, the proposed method can limit the number of elements considered in the original risk assessment;

(iv) A technique that combines Dempster Shafer Theory and the Analytic Hierarchy Process (namely, DST-AHP) is applied to the developed framework to assess the trustworthiness by a weighted average of the attributes in the framework: the AHP method is used to derive the weights of the attributes and the DST is used to account for the subjective uncertainty in the experts' judgments for the evaluation of the weights;

(v) A MHRA technique is developed based on Bayesian model averaging, to overcome the limitations of the current practice of risk aggregation that neglects the trustworthiness of the risk assessment of individual hazard groups;

(vi) The developed framework is applied to real case studies from the Nuclear Power Plants (NPP) industry.

Table of contents

List of Figures

List of Tables

Part (I): Thesis

Acronyms

ASME:	American Society of Mechanical Engineers
AHP:	Analytical Hierarchical Process
BBA:	Basic Belief Assignments
CMM:	Capability Maturity Model
CCS:	component cooling system
CDF:	Core Damage Frequency
DM:	Decision Making
DST:	Dempster Shafer Theory
DST-AHP:	Dempster Shafer Theory-Analytical Hierarchy Process
EDF:	Electricité De France
FT:	Fault Tree
LOCA:	Loss of Coolant Accidents
M&S:	Model and Simulation
MCDA:	Multi-Criteria Decision Analysis
MHRA:	Multi-Hazards Risk Aggregation
MSPM:	Multi-State Physics-based Model
MSM:	Multi-States Models
NPP:	Nuclear Power Plants
NRC:	Nuclear Regulatory Commission
NUSAP:	Numeral Unit Spread Assessment Pedigree
PBM:	Physical Based Models
PCMM:	Prediction Capability Maturity Model
PRA:	Probabilistic Risk Assessment
PSA:	Probability Safety Assessment
QRA:	Quantitative Risk Assessment
RHR:	Residual Heat Removal
RIDM:	Risk-Informed Decision-Making
SEI:	Software Engineering Institute
SoK:	Strength of Knowledge

Chapter 1 Introduction

This thesis addresses the issue of evaluating the trustworthiness of risk assessment for the purpose of Decision-Making (DM) and Multi-Hazards Risk Aggregation (MHRA). To contextualize the research of this thesis, Sect 1.1 revises the concept of risk assessment and introduces the problem of MHRA. Some open issues are identified. Sect. 1.2 reviews the literature on these open issues and Sect. 1.3 states the technical issues and the motivation of the thesis. Sect. 1.4 presents the structure of the thesis, in connection to the appended scientific papers. Finally, the scientific contributions of the thesis are discussed in Sect. 1.5.

1.1. Risk assessment

Probability Risk Assessment (PRA) is widely applied in various industries to quantify risk, e.g., nuclear, aerospace, chemical, etc. The results of a PRA are used to support safety-related decisions [1]. In PRA, models are used to represent systems and processes, and provide estimates of risk metrics [2]. These models are built on a set of assumptions that are translated into quantitative assessments through mathematical models and computer codes [3], [4], [5]. The risk assessment models need to balance between the accurate representation of the phenomena in the system or process, and the definition of the proper level of detail of their description [3].

The PRA results, then, depend on different modeling factors such as: the strength of the background knowledge and information available on the systems and processes [6], [7], [8], [3], [9], the validity of the assumptions made [10], [11], [7], the phenomenological understanding of the systems and processes [6], the validity of the models used [12], [9], the level of details of the descriptions, etc. [13]. The confidence that the decision maker can put on the results of a PRA depends on these factors. Communicating the solidity and strength of these factors in the risk descriptions obtained from PRA is very important for informing the DM. For example, if a decision maker is to choose between two risk reduction measures, he/she would choose the one leading to lower risk, provided that it is physically and economically feasible; however, he/she might reconsider the decision if it is known that the risk results supporting the chosen reduction measure are less trustworthy than for the other.

PRA models characterize risk by probabilistic indexes [6], where numerical values are calculated on the basis of a “model of the world” [14] developed on the basis of the available knowledge on the problem. Then, the Strength of Knowledge (SoK) supporting the risk assessment must be considered [6], [15], [16],

[17], [18].

Also, the risk considered can some from multiple sources. When the system of interest is subject to multiple hazards (e.g., a NPP exposed to the risk from internal events, external flooding, fires, etc.), MHRA must be performed to combine the knowledge on the risk from the different contributing sources [19]. This is done by developing different PRA models for the different contributors, with different degrees of trustworthiness [19], [20] and, then, aggregating them. The current practice of MHRA consists of a simple arithmetic summation of the risk values obtained with the PRA models of the risk contributors [19], without considering their different degrees of trustworthiness. However, a simple summation of the risk estimates without accounting for the degree of trustworthiness may lead to results that are misleading for the DM [19].

In summary:

(i) the risk description should be extended to cover also the factors affecting the trustworthiness of the risk assessment;

(ii) the MHRA should not be limited to a simple arithmetic summation over the risk contributors, but should also consider the level of confidence for DM [19].

1.2. Literature review

Risk assessment methods and supporting tools for complementing the description and communication of risk are reviewed in Sect. 1.2.1; MHRA tools for aggregating the risk indexes of different hazard groups are reviewed and discussed in Sect. 1.2.2.

1.2.1. Risk characterization

New perspectives have been recently proposed to generalize the probabilistic formulation of risk by adopting uncertainty instead of probability (which is a specific way of quantifying uncertainty). In [8], risk is described in terms of events, consequences, uncertainty (A, C, U) and a conceptual structure is presented for linking to it the elements of a Data-Information-Knowledge-Wisdom hierarchy. In [6], uncertainty is regarded as the main component of risk and probability as an epistemic-based expression of uncertainty [6], so that the representation of risk is broadened to cover the events, consequences, predictions, uncertainty, probability, sensitivity and knowledge represented by A,C,C*,U,P,S and K respectively. A simple practical method is proposed to identify uncertainty factors as inter-alia assumptions and presuppositions (solidity of assumptions), historical field data (availability of reliable data), understanding of phenomena and agreement among experts. In [7], the available knowledge is recognized

as a key factor for the trustworthiness of the risk assessment outcomes and a framework is proposed for evaluating it.

The assumptions made in PRA are also considered a key factor for the use of risk assessment to inform DM. An application of Numeral Unit Spread Assessment Pedigree (NUSAP) was proposed for analyzing the strength, importance and potential value-ladenness of assumptions through a pedigree diagram. The pedigree diagram covers seven criteria for evaluating the quality of assumptions: (i) plausibility; (ii) inter-subjectivity peers; (iii) inter-subjectivity stakeholders; (iv) choice space; (v) influence situational limitations; (vi) sensitivity to view and interests of the analyst (vii) and influence on results [10], [21], [22], [23]. Value ladenness is considered an independent variable that affects the quality of assumption in [7] and evaluated using seven main criteria (i) personal knowledge; (ii) sources of information; (iii) non-biasedness; (iv) relative independence; (v) past experience; (vi) performance measure; (vii) agreement among peers [7], [24].

In [17], the "assumptions deviation risk" tis introduced to reflect the criticality of assumptions. For assessing this, the main assumptions on which the analysis is based are first identified and, then, converted into a set of uncertainty factors obtained by evaluating: (i) the degree of expected deviation of the assumptions from reality and the consequences, (ii) a measure of uncertainty of the deviation and consequences, (iii) the knowledge on which the assumptions are based. Finally, a score is assigned to each deviation to reflect the risk related to the deviation of the assumptions and their implication on safety.

In [11], four approaches for treating uncertain assumptions are summarized: (i) law of total expectation; (ii) interval probability; (iii) crude SoK and sensitivity categorization; (iv) assumption deviation risk [11]. For the latter, the method proposed in [17], [25] is extended into a general and systematic framework for treating "uncertain" assumptions in risk assessment models. In this approach, an assumption is placed in one of six "settings", given the belief in the deviation from the assumption, the sensitivity of the risk index and its dependency on the assumption, and the SoK on which the assumptions are made. Guidance for the treatment of uncertainty related to the deviation of assumptions is given for each setting. The guidelines are based on the precept that with the increasing importance and criticality of an assumption, and the implication of its potential deviations, the effort exerted for characterizing its uncertainty should be increased.

An approach for integrating the "assumptions deviation risk" in PRA is presented in [26]. In this approach, the risk of assumption deviation is evaluated through five steps: (i) the safety objectives are first

defined; (ii) the critical assumptions on which risk assessment depends are identified; (i) the deviation scenarios required to violate the safety objectives are defined; (iv) the likelihood that such deviation could occur is assessed; (v) the SoK supporting the assessment is evaluated.

In [3], besides parametric uncertainty (epistemic uncertainty about the true values of the model parameters), the assumptions and approximations are identified as elements of model uncertainty to be accounted for by means of different approaches, including subjective and imprecise probabilities and semi-quantitative schemes.

In [27], uncertainty in model predictions arising from model parameters and the model structure is discussed. Two main attributes are introduced to define model uncertainty: model credibility and model applicability [28]. Model credibility refers to the quality of the model in estimating the unknown in its intended domain of application and is defined by a set of attributes related to the model-building process and utilization procedure (conceptualization and implementation, which are in turn broken down into other sub-attributes). On the other hand, model applicability represents the degree to which the model is suitable for the specific situation and problem (represented by the conceptualization and intended use function attributes) [28].

Some works can be found in the literature for evaluating the trustworthiness of a model and other related quantities. In [29], the trustworthiness of risk assessment models is evaluated through a hierarchical tree of different factors i.e., modeling fidelity, SoK, number of approximations, amount and quality of data, quality of assumptions, number of model parameters etc. In [30], the trust of the model is evaluated based on the level of its maturity, evaluated through four main criteria: (i) uncertainty; (ii) knowledge; (iii) conservatism; (iv) sensitivity.

Credibility and maturity of Model and Simulation (M&S) processes have also attracted attention. For example, in M&S and information systems, the Capability Maturity Model (CMM), developed by the Software Engineering Institute (SEI), has been developed to assess the maturity of a software development process in the light of its quality, reliability and trustworthiness, considering: representation and geometric fidelity, physics and material model fidelity, code and solution verification, model validation, uncertainty quantification, and sensitivity analysis [31]. In [9], a hierarchical framework has been developed to assess the maturity and prediction capability of a prognostic method for maintenance DM purposes. The hierarchical tree covers different attributes that are believed to affect the prognostic method prediction capability. In [12], a framework is proposed for assessing the credibility of M&S through eight criteria: (i)

verification; (ii) validation; (iii) input pedigree; (iv) results uncertainty (v) results robustness; (vi) use history; (vii) M&S management; (viii) people qualification. Finally, the quality of M&S is assured by two steps in the American Society of Mechanical Engineers (ASME) i.e., verification and validation [32]. Verification concerns the accuracy of the computational model in representing the conceptual and mathematical model, and validation is related to the accuracy of the model in representing reality [32].

Some open issues related to the evaluation of the trustworthiness of risk assessment outcomes are:

(i) most of the aforementioned works treat the factors contributing to trustworthiness, without integrating them in a comprehensive framework;

(ii) the evaluation of the SoK and model trustworthiness is done by directly scoring some intangible contributing factors, without breaking them into more tangible attributes, easier to evaluate in practice;

(iii) trustworthiness is not integrated in the results of risk assessment.

1.2.2. MHRA

Few works in the literature focus on MHRA and a relatively recent report by EPRI [19] indicates that current practice might not be appropriate for some DM contexts, due to the difference in the degrees of confidence on the risk contributors. The report also highlights some of the fundamental differences in the risk estimates from different hazard sources (e.g., maturity of the used tool and analysis, uncertainty level for each contributor). Then, it proposes a practical guidance for an integrated understanding of the risk to support DM, within the context of RG1.174 [33]; the USNRC regulatory guide on using PRA in RIDM (i.e., meet the current regulations, meet the defense in depth requirement etc.; see Figure 2 in [33] for more information). This is done by developing the relevant insights for each of the contributions to risk. Five main tasks are “iteratively” performed according to this guidance: (i) understand the role of PRA in supporting the decision; (ii) identify the main risk contributors and assess the baseline risk and evaluate the confidence in the assessment; (iii) evaluate relevant risk metrics and refine the PRA if needed; (iv) identify and characterize key sources of uncertainty; (v) document conclusions for integrated DM. No clear guidance, however, is provided on how to evaluate the level of trustworthiness in risk assessment.

An iterative method is proposed also in [34] for assessing different aspects of risk, aggregated from highly heterogeneous hazard groups, focusing on relative rather than absolute risk metrics. The method uses response surfaces that are based on arbitrary polynomial chaos expansion in combination with radar charts to visualize the overall risk and associated uncertainties. The response surface allows identifying

major contributors to the overall risk, individually or on aggregate basis for a very large number of input parameters. On the other hand, radar charts are used to visualize risk contributors of different nature and compare them to safety guidelines. However, the method does not address factors like model conservatism, biases, incompleteness, hidden model uncertainty (e.g., structural), etc. Also, radar charts do not really allow the aggregation of risk from different contributors. Instead, they only allow the relative comparison of the risk contributors (hazard groups) to a given threshold.

1.3. Technical issues and motivation of the thesis

The main objective of a risk assessment is to provide informative supports to DM [35], [36], [5], [3], [34]. Also, the current practice of MHRA is that of a simple arithmetic summation of the individual risk indexes, without considering the level of trustworthiness of the assessment of different risk contributors. With respect to these issues, the work presented in this thesis focuses on:

(1) the development of an integrated framework to evaluate the level of trustworthiness of a risk assessment, considering all contributing factors;

(2) the development of a MHRA framework that allows the integration of the level of trustworthiness of the risk assessment of the individual hazard groups in the aggregation process.

1.4. Structure of the thesis

Risk assessment is performed using models and performing analyses that are supported by background knowledge, including data, phenomenological understanding on the involved systems and process, etc. The quality of the assessment depends also on other factors like the quality of the assumptions made, the maturity of the analysis, the tools used, etc. In this thesis, these factors are included in an integrated framework for assessing the trustworthiness of the risk assessment. Trustworthiness is, then, integrated in the MHRA to support safety-related DM.

The research included in this thesis can be divided into three main parts, as shown in Figure 1.1. In the *first part* (Chapter 2), an integrated framework is developed for assessing the trustworthiness of risk assessment. Then, in the *second part* (Chapters 3-5), maturity of analysis, assumptions and SoK that support the risk assessment, are considered. Finally, in the *third part* (Chapter 6), the two previous parts are integrated in a complete framework for evaluating the trustworthiness of risk assessment, and a technique is developed based on the weighted posterior method for MHRA considering the level of trustworthiness.

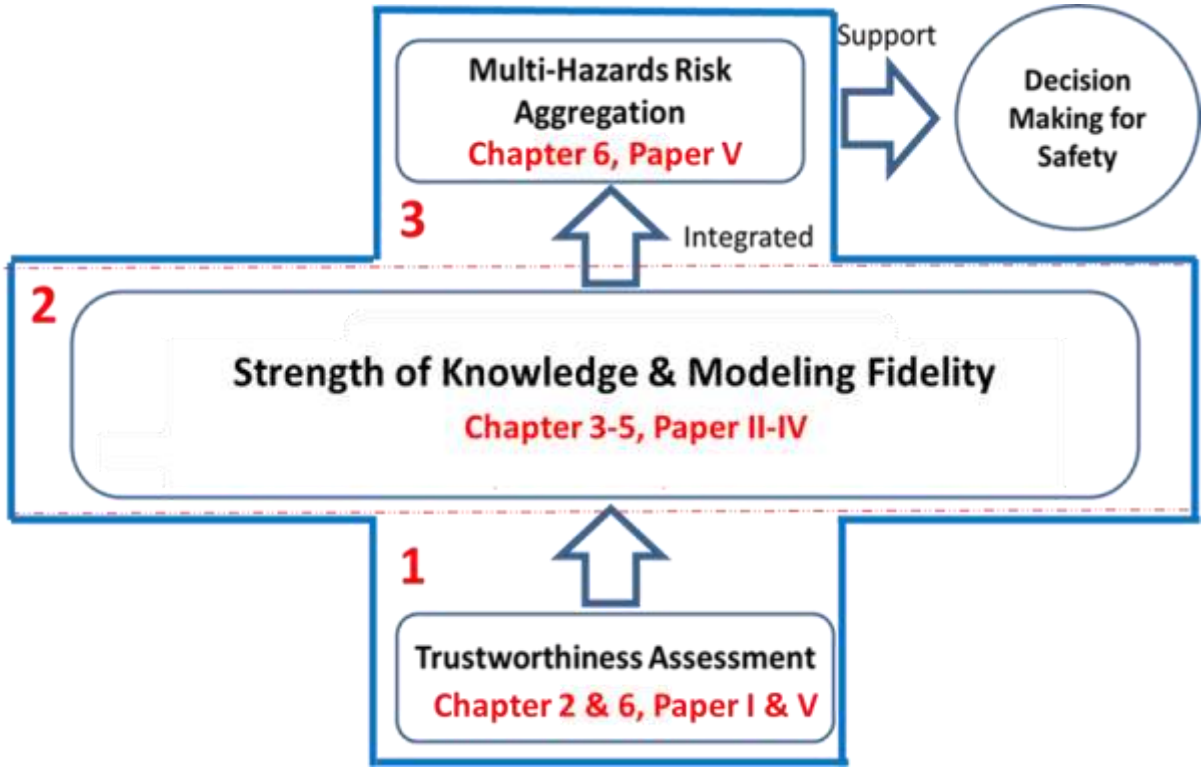


Figure 1.1 Conceptual scheme of the thesis work

In chapter 2 and the appended paper I, we discuss trustworthiness in risk assessment and propose a four-levels, top-down, hierarchical tree to identify the main attributes and criteria that affect the level of trustworthiness of the models used in probabilistic risk assessment. The level of trustworthiness is decomposed into two attributes (Level 2), three sub-attributes (Level 3), one "leaf" attribute (Level 3) and seven basic "leaf" sub-attributes (Level 4). On the basis of this hierarchical decomposition, a bottom-up, quantitative approach is employed for the assessment of model trustworthiness, using tangible information and data available for the basic "leaf" sub-attributes (Level 4). Analytical Hierarchical Process (AHP) [37] is adopted for evaluating and aggregating the sub-attributes.

In chapter 3 and the appended paper II, we elaborate on some of the main contributing factors to the trustworthiness related to the maturities of risk assessment. In particular, we propose a hierarchical framework to evaluate the level of maturity of risk contributors in the light of DM. The framework consists of four attributes that are believed to affect the level of maturity of risk analysis, i.e., uncertainty, conservatism, knowledge and sensitivity. The knowledge attribute is, in turn, decomposed into five sub-attributes i.e., availability of data, consistency of data, data reliability, experience, and value-ladenness. AHP is again adopted for the application of the framework to assess the level of maturity. A reduced-order model technique is used to enable the application of the framework on a real problem. Then, the maturity

level is integrated in MHRA for a two-dimensional risk aggregation method. Scoring protocols for evaluating the attribute have been prepared to simplify the application of the framework and to reduce the subjectivity of the assessors. Finally, a numerical case study for the MHRA of a real NPP is carried out to show applicability.

In chapter 4 and the appended paper III, we elaborate on the factors contributing to trustworthiness that are related to the assumptions in risk assessment models, to understand their implication on the trustworthiness in risk assessment models. In particular, we develop an extended framework for evaluating the risks that deviations from the assumptions made lead to a reduction of the safety margins. We extend the framework in [26] to cover also the risk of deviations from conservative assumptions and other contexts of DM and, then, introduce decision flow diagrams for the quantitative evaluation of the assumption deviation risks. Finally, we apply the framework to a real case study from the nuclear industry.

In chapter 5 and the appended paper IV, we focus on the importance and the influence of background knowledge on the trustworthiness of risk analysis and zoom in on this particular attribute in order to provide a comprehensive evaluation approach. In particular, we develop a new quantitative method to assess the SoK of a risk assessment. A hierarchical framework is first developed to conceptually represent the SoK in terms of three attributes (assumptions, data, phenomenological understanding), which are further decomposed in sub-attributes and "leaf" attributes to facilitate their assessment in practice. The hierarchical framework is, then, quantified in a top-down bottom-up fashion for assessing the SoK. In the top-down phase, a reduced-order risk model is constructed to limit the complexity and number of basic elements considered in the SoK assessment. In the bottom-up phase, the SoK of each basic element in the reduced-order risk model is assessed based on predefined scoring guidelines and, then, aggregated to obtain the SoK for the whole risk assessment model. The aggregation is done using a weighted average of the basic events' SoK, where the weights are determined by AHP. The developed methods are applied to a real-world case study, where the SoK of the PRA models of a NPP is assessed for two hazard groups, i.e. external flooding and internal events.

Finally, in chapter 6 and the appended paper V, we integrate the previous efforts to develop a more complete and comprehensive framework for evaluating the trustworthiness of risk assessment, and, then, develop a new method for MHRA considering the level of trustworthiness. In particular, a hierarchical framework is first developed for evaluating the trustworthiness of risk assessment. The framework is based on two main attributes (criteria) i.e., the SoK and modeling fidelity, which are further decomposed into

sub-attributes and leaf attributes on different levels. The trustworthiness is calculated using a weighted average of the leaf attributes, where the weights are calculated using the Dempster Shafer Theory-Analytical Hierarchy Process (DST-AHP). A technique is, then, developed to update the model output risk estimates considering the level of trustworthiness and, finally, aggregate the risks from different hazard groups. The developed framework is, then, applied to a real case study of two hazard groups in a NPP.

1.5. Contributions

The scientific contributions of this thesis are:

(i) Factors contributing to the trustworthiness of risk assessment outcomes are identified and their criticalities are analyzed under different frameworks, to understand their influence on the risk results (Chapter 2-6, papers I-IV);

(ii) An integrated hierarchical framework is developed for assessing the trustworthiness of risk analysis, based on the identified factors and related attributes (Chapter 6, paper V);

(iii) A reduced order model-based method is proposed to efficiently evaluate the trustworthiness of risk assessment in practice. Through the reduced-order model, the proposed method can limit the number of elements considered in the original risk assessment (Chapters 3 and 5, Papers II and IV);

(iv) A technique that combines Dempster Shafer Theory and the Analytic Hierarchy Process (namely, DST-AHP) is applied to the developed framework to assess the trustworthiness by a weighted average of the attributes in the framework: the AHP method is used to derive the weights of the attributes and the DST is used to account for the subjective uncertainty in the experts' judgments for the evaluation of the weights (Chapter 6, Paper V);

(v) A MHRA technique is developed based on Bayesian model averaging, to overcome the limitations of the current practice of risk aggregation that neglects the trustworthiness of the risk assessment of individual hazard groups (Chapter 6, Paper V);

(vi) The developed framework is applied to real case studies from the Nuclear Power Plants (NPP) industry (Chapter 6, Paper V).

The contents in the thesis are based on a series of submitted papers. Table 1.1 shows how does each chapter correspond to the appended papers and the contributions.

Table 1.1 Structure of the thesis

Chapters	Associated papers	Contributions
Chapter 2. Assessing the trustworthiness of risk assessment models	**I**	**i, ii**
Chapter 3. Risk analysis model maturity index for Multi-Hazards Risk Aggregation purposes	**II**	**i, iii**
Chapter 4. Assumptions in risk assessment models and the criticality of their deviations within the context of decision making	**III**	**i**
Chapter 5. Strength of knowledge supporting risk analysis: assessment framework	**IV**	**i**
Chapter 6. Framework for multi-hazards risk aggregation considering trustworthiness	**V**	**ii, iv, v**

Chapter 2 Assessing the trustworthiness of risk assessment models

Risk assessments rely on the use of complex models to represent systems and processes, and provide predictions of safety performance metrics [2]. Since the fundamental value of a risk assessment lies in providing informative support to (high-consequence) decision making, the importance placed on Modeling and Simulation (M&S) is very high within a risk assessment context. Accordingly, the confidence that can be put on the results of a risk assessment is fundamental for DM. Therefore, quantitative measures that relate to the credibility and trustworthiness of risk assessment outcomes must be provided to be used for DM purposes.

Within this context, the objective of this chapter is to survey the factors that affect the credibility and trustworthiness of risk assessment models, and organize them within a "preliminary" assessment framework. A review of the approaches proposed in the literature to assess the trustworthiness and credibility of a model is presented in Sect. 2.1. In Sect. 2.2, a hierarchical tree-based framework for assessing model trustworthiness is presented. In Sect 2.3, we review and explain the Analytic Hierarchy Process (AHP) for assessing trustworthiness within the developed framework. In Sect. 2.4, the framework is applied to a real case study concerning the RHR system of a NPP. Finally, Sect. 2.5 discusses the results and draws conclusions.

2.1. State of the art

Few methods have been proposed to assess the credibility and trustworthiness associated with engineering model predictions. In the literature, the trustworthiness of a method or a process is often measured in terms of its maturity. The model maturity was previously used to assess the maturity of a function of an information system [31],[38],[9]. Later, the SEI developed a framework known as the CMM to assess the maturity of a software development process, in the light of its quality, reliability and trustworthiness [39]. Recently, the CMM model has been extended to what so-called a Prediction Capability Maturity Model (PCMM) evaluate and assess the maturity of modeling and simulation efforts [31]. Other examples of maturity assessment approaches have been developed in different domains, such

as data maturity assessment, enterprise risk management and hospital information system [9]. In [40] and [9] a hierarchical framework based on the AHP has been developed to assess the maturity and prediction capability of a prognostic method for maintenance DM purposes. Finally, a framework for assessing the credibility of M&S is proposed by [12]. In this framework, three main groups of criteria are used to assess the credibility of M&S (i) M&S development including; (ii) M&S operations (iii) supporting evidence. These are in turn cover verification, validation, input pedigree, results uncertainty, results robustness, use history, M&S management, and people qualifications. However, most of the aforementioned works are not complete in the sense of evaluating the trustworthiness of risk assessment models Also, they do not present a rigorous evaluation protocols for the attributes and criteria. Instead, the evaluation of criteria is done by directly scoring the some intangible contributing factors, which is hard to apply in practice.

2.2. Hierarchical tree for model trustworthiness characterization: abstraction and decomposition

Many factors (attributes) affect the trustworthiness and credibility of analyses and models (for risk assessment in particular), and several studies and literature reviews have been made in order to identify them. Some of these are summarized as follows: (i) phenomenological understanding of the problem; (ii) availability of reliable data; (iii) reasonability of the assumptions; (iv) agreement among the experts; (v) level of detail in the description of the phenomena and processes of interest; (vi) accuracy and precision in the estimation of the values of the model's parameters; (vii) level of conservatism; (viii) amount of uncertainty and others (see e.g., [6], [11], [8], [41]; [1], [3], [9], [31], [36], [19], [7]). However, these attributes (criteria) are not tangible and cannot be measured directly: as a consequence, other sub-attributes must be identified, which can be measured directly or subjectively scored. To this aim, we propose a method for model trustworthiness characterization and decomposition, which is based on the hierarchy tree shown in Figure 2.1. See the appended paper I for the detailed discussions.

As mentioned above, many factors can be found in the literature that characterize the level of trustworthiness. Those factors can be categorized into two main groups: (i) "strength of knowledge"; (ii) "modeling fidelity", which embody the ability of a model of representing the reality and the degree of implementing correctly the model. In the "strength of knowledge", among the four sub-elements proposed in [6], two were found to be more relevant to the context of interest, i.e., data and assumptions. In the modeling fidelity, it is argued that including more details about a problem is more representative and realistic, and hence more trustworthy. On the other hand, implementing the model correctly from a pure

trustworthiness point of view, without considering a costs-benefits reasoning, requires avoiding approximation: the less the approximations, the better the trustworthiness is. In accordance, a hierarchical tree for models' trustworthiness is proposed in Figure 2.1.

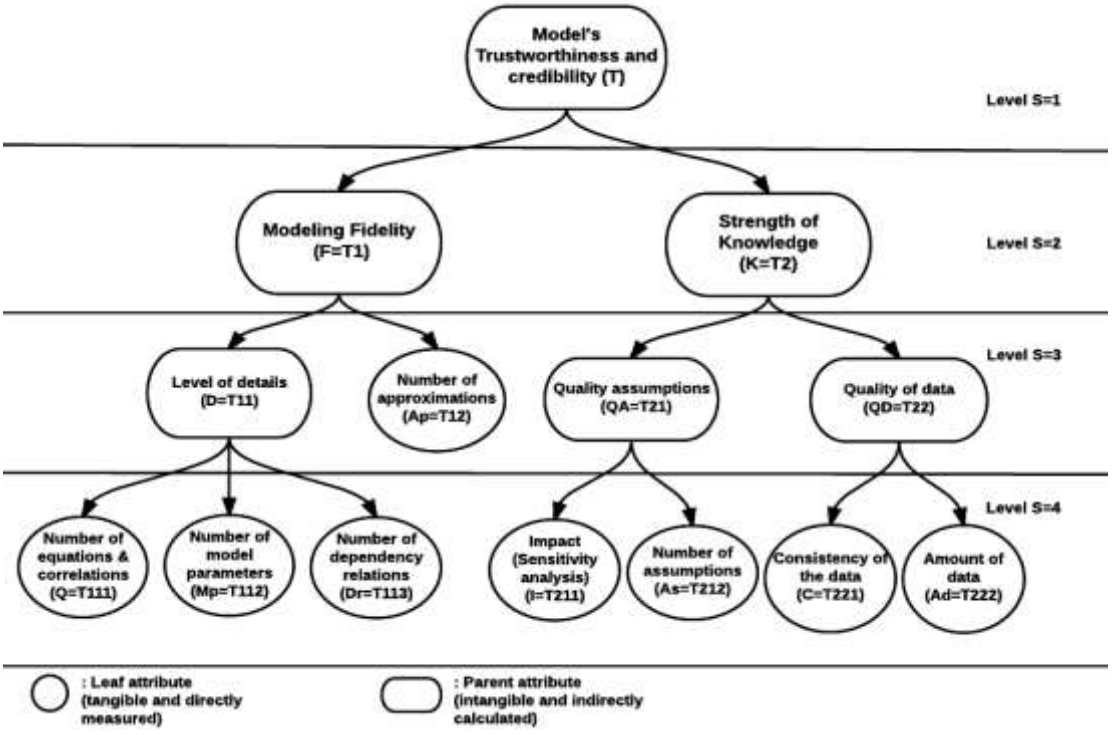


Figure 2.1 A hierarchical tree-based framework for the trustworthiness of mathematical models

The model trustworthiness, represented by T (Level 1), is characterized by two attributes: modeling fidelity, represented by $F = T_1$ and strength of knowledge, represented by $K = T_2$ (Level 2). The modeling fidelity ($F = T_1$), measures the adequacy of the model representation of the phenomenon and the level of detail adopted in the model description (referred to as modeling validity in some literatures [42]).On the other hand, the strength of knowledge ($K = T_2$) measures how solid the assumptions, data and information (which the model relies on) are [6]. These two attributes are in turn decomposed into sub-attributes (Level 3). In particular, the modeling fidelity $F = T_1$ is defined by the level of detail, represented by $D = T_{11}$ (Level 3) and by the number of approximations, represented by $Ap = T_{12}$. Concerning the strength of knowledge $K = T_2$, among the four sub-attributes proposed in [43], i.e., the solidity of assumptions, the availability of reliable data, understanding of phenomena, and agreement among experts, two are found to be more relevant to the context indeed, i.e. data and assumptions. Thus, attribute $K = T_2$ is here defined by the quality of assumptions represented by $QA = T_{21}$ and by the quality of data represented by $QD = T_{22}$. Note that the number of approximations $Ap = T_{12}$ is considered as a basic

attribute, since it can be measured directly: thus, it is not further broken down into other sub-attributes. The other three attributes of Level 3 are instead broken down into more basic "leaf" attributes that can be measured directly. In particular, the level of detail $\mathrm{D} = \mathrm{T}_{11}$ is characterized in terms of the number of equations and correlations, namely $\mathrm{Q} = \mathrm{T}_{111}$, the number of model parameters, namely $\mathrm{Mp} = \mathrm{T}_{112}$, and the number of dependency relations included, namely $\mathrm{Dr} = \mathrm{T}_{113}$. The overall quality of the assumptions $\mathrm{QA} = \mathrm{T}_{21}$ is measured by the number of assumptions made $\mathrm{As} = \mathrm{T}_{212}$, and by their impact $\mathrm{I} = \mathrm{T}_{212}$ (which can be assessed, e.g., by sensitivity analysis). Finally, the quality of the data $\mathrm{QD} = \mathrm{T}_{22}$ is described in terms of the amount of data available, namely $\mathrm{Ad} = \mathrm{T}_{221}$ and by the consistency of the data itself, namely $\mathrm{C} = \mathrm{T}_{222}$. Precise definitions of the attributes are given in Table 2.1 for the sake of clarity.

Table 2.1 Definition of the attributes used to characterize the model trustworthiness

Attribute	Definition
Modeling fidelity $F = T_1$	Measures how close the model is to reality, i.e., the adequacy of the representation of the phenomena and processes of interest: the higher the modeling fidelity, the higher the trustworthiness of the model.
Strength of knowledge $K = T_2$	Represents the level of understanding of the phenomena and the solidity of the assumptions, data and information, which the model relies on: the higher the strength of knowledge, the higher the trustworthiness of the model.
Level of detail $D = T_{11}$	Measures the level of sophistication of the analysis by quantifying to which level the "elements" and aspects of the phenomenon, process or system of interest are taken into account in the model: the higher the level of detail, the higher the trustworthiness of the model.
Number of approximations $Ap = T_{12}$	Measures the number of approximations that the analyst introduces in order to facilitate the analysis: it affects the modeling fidelity. The lower the number of model approximations the higher the modeling fidelity.
Quality of assumptions $QA = T_{21}$	In some studies, experts are obliged to formulate some assumptions, which might be due to the lack of data and information, to the complexity of the problem or to lack of phenomenological understanding. The quality of those assumptions is an indication of the strength of knowledge: the higher the quality of the assumptions, the higher the trustworthiness of the model.
Quality of data $QD = T_{22}$	Represents the availability of sufficient, accurate and consistent background data with respect to the purposes of the analysis: the higher the quality of the data, the higher the trustworthiness of the model.
Number of equations and correlations $Q = T_{111}$	The number of equations and correlations used in modeling is an indication of the level of detail, hence of the modeling fidelity: the higher the number of equations and correlations, the higher the trustworthiness of the model.
Number of model parameters $Mp = T_{112}$	The number of parameters introduced in the model is a measure of the level of detail (e.g., the number of components transition rates represents the level of discretization adopted to describe the failure process of a component or a system): the higher the number of model parameters, the higher the

trustworthiness of the model.

Number of dependency relations $Dr = T_{113}$	The larger the number of dependency relations that are taken into account, the more detailed and trustworthy the model.
Number of assumptions $As = T_{211}$	The larger the number of high-quality assumptions, the higher the trustworthiness of the model.
Impact of assumptions $I = T_{212}$	It quantifies how much assumptions can affect the model results (and it can be assessed by sensitivity analysis). The higher the impact of the assumptions, the lower the trustworthiness of the model.
Consistency of data $C = T_{221}$	It is an indication of how suitable and representative the data are for a specific process or system. The consistency of data relies on the sources of the data. For example, if we are collecting data about the failure of a safety system's pump from different power plants, we should first understand whether the power plants are of the same type, whether the plants work at the same power level and whether the pumps have the same work function and capacity.
Amount of data $Ad = T_{222}$	The higher the amount of data available, the stronger the knowledge. For example, the number of years of experience of a particular component in a plant can be sometimes considered an indication of the amount of data available. In any domain, a higher number of years' experience means a higher number of scenarios covered and hence a larger amount of data. The higher the amount of data, the higher the trustworthiness of the model.

It should be noted that the approach proposed might not be comprehensive and complete. For example, an increase in the number of parameters of a model, on one side, increases the level of details that the model is capable to capture but, on the other side, may leave room for additional errors and uncertainties in its estimated parameters (which are not included in the present formulation). As specified before, the constituting attributes have been selected on the basis of an accurate and critical literature review of works treating the subject. Also, guidelines have been developed to provide scoring protocols that facilitate the evaluation process. These guidelines help in overcoming the problem of evaluating some attribute that have contrasting effect on model trustworthiness, e.g., number of approximations (a lower score is given for a higher number of approximations). These guidelines have been developed on the basis of the experience and knowledge of Electricité De France (EDF) experts (see Appendix A in the appended paper I). So, the contribution here is considered as a first attempt of a systematic framework to address the evaluation of model trustworthiness and to give a structure to organized expert judgments on this. The framework is refined in Chapter 6 for a complete description and assessment of trustworthiness.

2.3. Analytical hierarchical process (AHP) for model trustworthiness quantification

Given the hierarchical tree in Figure 2.1, the assessment of model trustworthiness is carried out within a Multi-Criteria Decision Analysis (MCDA) framework [44]; [45]. In this setting, we suppose that a system, process or phenomenon of interest for a risk assessment can be represented by different mathematical models of possibly different complexity and level of detail, $M1, M2, .. M_l, .., M_n$. The task (i.e., the MCDA problem at hand) is to rank these alternative models with respect to their trustworthiness, in relation to the particular risk assessment problem of interest to support MCDA. In the present chapter, the Analytical Hierarchy Process (AHP) proposed by [46] is adopted to this aim.

2.3.1. Introduction to analytical hierarchical process

AHP is a MCDM method that is known for its capability of considering both quantitative and qualitative evaluations of attributes and factors [47] and it can be helpful in group-decision-making [48]. This method is usually used for decreasing the complexity of comparison process for decision-making purposes, as it allows comparing only two criteria (or alternatives) at a time and then computing the "overall" relative importance of a criterion in a group of criteria. In addition, it allows gauging and enhancing the rationality and consistency of the expert's evaluation for the criteria by measuring the consistency of the pairwise comparison matrices. Pairwise comparison matrices are first constructed in AHP for assessing the relative importance of criteria. Then, the local relative importance of different alternatives are compared with respect to the criteria hierarchically. Decisions are made based on the overall all relative importance of each alternative [49].

In this approach, the top goal, i.e., the decision problem considered (in this case, ranking the model trustworthiness), is placed at the first level of the hierarchy and, then, decomposed into several sub-attributes distributed over different levels according to their degree of tangibility. Finally, the bottom level in the hierarchal tree-based AHP model contains the different alternatives that need to be evaluated with respect to the top goal (i.e., in this case the level of trustworthiness) [48], [9]. Through pairwise comparisons among the elements and the attributes of the same level, the alternative solutions, i.e., models, can be ranked with respect to the decision problem in the top level (i.e., the model trustworthiness) [48], [50].

The AHP model for model trustworthiness assessment is represented in Figure 2.1. The first step required to assess the model trustworthiness by AHP is the determination of the so-called inter-level priorities (in practice, weights that represent the importance of attributes in the same level relative to their parent attribute) for each attribute, sub-attribute, basic "leaf" sub-attribute and alternative solution i.e.,

$W(T_i)$, $W(T_{ij}), W(T_{ijk})$, and $W(M_l, T_{ijk})$, respectively. Notice that in practice, each weight represents the relative contribution of an attribute of a given level to the corresponding "parent" attribute of the upper level: for example, weight $\mathrm{W}(T_{ijk})$ quantifies the contribution of basic "leaf" sub-attribute T_{ijk} (of Level 4) in the representation and definition of sub-attribute T_{ij} (of Level 3); instead, weight $W(M_l, T_{ijk})$ is the weight of the $l-th$ model with respect to the basic "leaf" sub-attribute T_{ijk}.

The weights $W(T_i)$, $W(T_{ij})$ and $W(T_{ijk})$ are calculated using pairwise comparison matrices: in particular, one pairwise comparison matrix is constructed for the attributes at the second level $S = 2$, one is constructed for each "set" of sub-attributes at level $S = 3$ that fall under the same "parent" attribute in the upper level $S = 2$, and one is constructed for each "set" of basic "leaf" attributes at level $S = 4$ that fall under the same "parent" sub-attribute in the upper level $S = 3$. The comparison matrix is a $(n \times n)$ square matrix, to be filled by experts, where *n* is the number of elements being compared. Attributes in each level are compared to each other with respect to their contribution in defining their "parent" attribute in the upper level. For example, a (3×3) matrix is constructed to compare the basic sub-attributes $Q = T_{111}$, $Mp = T_{112}$ and $Dr = T_{113}$ (Level 4), with respect to their "parent" sub-attribute $D = T_{11}$ (Level 3). Typically, experts use a scale from 1 to 9 to evaluate the strength (i.e., the contribution) of each criteria with respect to the other; for example, the scale suggested by Saaty [48] used to carry out a qualitative comparison between two attributes A and B, is the following:

1: A and B are equally important,

2: A is slightly more important than B,

3: A is moderately more important than B,

4: A is moderately-plus more important than B,

5: A is strongly more important than B,

6: A is strongly-plus more important than B,

7: A is very strongly more important than B,

9: A is extremely more important than B.

Another possibility is to use the *"generalized balanced scale"*, which is recommended due to its ability to overcome the problem of uneven dispersion of the local weights that could lead to inaccurate estimates. Please refer to appended paper 1 for more details about the balanced scale.

A pairwise comparison matrix is made for each group of attributes in the same level (say, S) sharing the same parent attribute in the upper level (S-1). Each expert is asked to fill individually the pairwise

comparison matrices as illustrated above. For each matrix, the weight of each attribute can, then, be determined by solving the eigenvector problem and normalizing the principal eigenvectors (for details, see [48], [46], [49]). A good approximation for calculating simply the eigenvector is by multiplying the elements in each row and then to take the n-th root of the product (n is the matrix size). The output of the row is eventually, normalized with the other row's outputs.

It should be noted that the consistency of the pairwise comparison matrix should be checked by calculating the consistency ratio (CR):

$$CR = \frac{CI}{RI}, \tag{2.1}$$

where RI represents the consistency index of a randomly generated matrix and its value can be taken from Table 1 in [51], and CI is the consistency index which is calculated by Eq. (2.2):

$$CI = \frac{\lambda_{max} - n}{n-1}, \tag{2.2}$$

where λ_{max} is the maximum eigenvalue and n is the order of the matrix and represents the number of attributes being compared [48], [24]. Saaty's acceptance criteria of consistency is adopted [48]: when $CR < 0.1$, the comparison matrix is consistent, otherwise it is not and the experts are demanded to revise their evaluations [24] [52], [51]. After checking the consistency of the matrices and obtaining the weights of the attributes from each expert. The final weight of each attribute is calculated by averaging the weights obtained from the experts. Notice that the weights obtained should be normalized to sum to 1 at each hierarchy.

An illustration example on how to apply the AHP for determining the weights of is given below. Let's take again the level of details $D = T_{11}$ at Level 3 as an example. The level of details has three daughter attributes at Level 4: the number of equations and correlations $Q = T_{111}$, the number of model parameters $Mp = T_{112}$, and the number of dependency relations $Dr = T_{113}$ (Level 4). A 3×3 pairwise comparison matrix is constructed to compare the basic sub-attributes. The experts are then asked to fill the pairwise comparison matrices in Table 2.2, in order to evaluate the importance of each attribute (criteria). The attributes relative importances with respect to the parent attribute (level of detail) have been evaluated using the 1-9 scaling.

The first step is to evaluate the consistency of the matrix. By solving the eigenvector problem, the maximum eigenvalue is found to be $\lambda_{max} = 3$. From Eqs. (2.1) and (2.2), the consistency ratio for this matrix is $CR = 0$, since the order of the matrix equals to maximum eigenvalue $\lambda_{max} = n = 3$. This

means that the matrix is consistent. Now for determining the weights, let's adopt the approximation illustrated previously. The 3rd root of the multiplication of the elements in each row is found and then the results are normalized to obtain the weights. For example, the relative importance of the first row is calculated as the following:

$$\sqrt[3]{1 \times 3 \times 1} = 1.44$$

Then it is normalized to 0.449 as illustrated in Table 2.2. Note that the weights of the three attributes in the example sum to one: $\sum_{k=1}^{3} W_{11k} = 1$.

Table 2.2 Pairwise comparison matrix for level of detail daughter attributes

	Q	Mp	Dr	Relative importance	Normalized weight
Q	1	3	1	1.44	0.449
Mp	1/3	1	1/3	0.33	0.102
Dr	1	3	1	1.44	0.449

2.3.2. Model trustworthiness quantification using AHP

For the tangible basic leaf sub-attributes T_{ijk}, a quantitative evaluation $T_{M_l,T_{ijk}}$ can be given directly if they are quantitative in nature. If the basic leaf sub-attributes are not quantitative in nature, the scaling system explained above (i.e., scores from 1 to 9) can be adopted to provide a (semi-quantitative) relative evaluation of the leaf attributes T_{ijk} with respect to the risk models M_l available (guidelines are provided in Appendix A in the appended paper I for relatively evaluating the basic leaf sub-attributes). Also, if the attribute is not the larger the better with respect to the trustworthiness, the scaling system provided in the guidelines needs to be adopted. For example, the larger the number of approximation, the worst the trustworthiness is. Therefore, this attribute needs to be evaluated given the guidelines provided in the Appendices in the appended paper I.

The corresponding inter-level weights $W(M_l, T_{ijk})$ can, then, be obtained as $\frac{T_{M_l,T_{ijk}}}{\sum_{l=1}^{n} T_{M_l,T_{ijk}}}$. The weights $W(M_l, T_{ijk})$ are normalized so that $\sum_{l=1}^{n} W(M_l, T_{ijk}) = 1$, where n is the number of models.

Finally, the normalized trustworthiness $T(M_l)$ of a model M_l is evaluated using a weighted average of the leaf attributes, as indicated in Eq. (2.3):

$$T(M_l) = \sum_{i=1}^{n_T} \sum_{j=1}^{n_{T_i}} \sum_{k=1}^{n_{T_{ij}}} W(T_i) \cdot W(T_{ij}) \cdot W(T_{ijk}) \cdot \frac{T_{M_l,T_{ijk}}}{\sum_{l=1}^{n} T_{M_l,T_{ijk}}} 2 \quad (2.3)$$

where $T_{M_l,T_{ijk}}$ is the numerical value that the basic "leaf" sub-attribute $T_{T_{ijk}}$ takes with respect to model M_l, (for

example, for attributes $Q = T_{111}$ variable $T_{M_l,T_{111}}$ equals the number of equations and correlations contained in M_l), n is the number of models to be compared, n_T, n_{T_i}, and $n_{T_{ij}}$ are defined above.

After obtaining the weight for each criterion with respect to the corresponding upper-level criteria, a "global" weighting for each criterion with respect to the top goal *T* can also be obtained by multiplying its weight by the weights of its upper parent elements in each level: for example, the "global" weight of basic "leaf" sub-attribute T_{ijk} with respect to the "top" attribute (goal) T is given by $W(T_{ijk}) \cdot W(T_{ij}) \cdot W(T_i) = W_{global}(T_{ijk})$. For example, in the hierarchical tree Figure 2.1, the "global weighting" of the "consistency of data" (denoted by T_{221}) with respect to level of trustworthiness is obtained by multiplying its weight by the weight of quality of data (denoted by T_{22}) by the weight of strength of knowledge (denoted by T_2): $W_{global}(T_{221}) = W(T_{221}) \cdot W(T_{22}) \cdot W(T_2)$. The trustworthiness $T(M_l)$ can then be expressed directly as a function of the "global" weights of the leaf attributes with respect to the top goal *T*:

$$T(M_l) = \sum_{i=1}^{n_T} \sum_{j=1}^{n_{T_i}} \sum_{k=1}^{n_{T_{ij}}} W_{global}(T_{ijk}) \frac{T_{M_l,T_{ijk}}}{\sum_{l=1}^{n} T_{M_l,T_{ijk}}} \quad (2.4)$$

In addition, the enumeration of some model leaf attributes (e.g., approximations, assumptions, formulas…) may be an "artifact" of presentation or interpretation, in absence of a protocol rigorously constructed to this aim. On the other hand, the following aspects should be considered. First, such a type of evaluation has been already used for evaluating some attributes in some relevant models e.g., evaluation of phenomenological understanding, availability of reliable data, reasonability of assumptions and agreement among peers, demonstrating the feasibility [6]. Second, the issue of enumerating model assumptions and evaluating their quality have already been treated in several papers: see, e.g., [17], [53]. Then, most importantly, notice that the "direct enumeration" is not the only way to provide numerical values $T_{M_l,T_{ijk}}$ for the basic "leaf" attributes $T_{T_{ijk}}$ with respect to the model M_l. As mentioned above, if the analyst does not feel confident in evaluating the assumptions, formulas and correlations quantitatively, he/she may resort to semi-quantitative scale (e.g., scores from 1 to 9), in order to provide a relative evaluation of a "leaf" attribute $T_{T_{ijk}}$ with respect to the different risk models M_l's available (see for example the enumerating protocols in Appendix A of the appended paper I, based on technical reports and experts' feedback).

2.4. Application

In this section, the hierarchical tree-based framework is applied to a case study concerning the modeling of the Residual Heat Removal (RHR) system of a NPP. In Sect. 2.4.1, the system is described; in

Sect. 2.4.2, the characteristics of the two models used to represent the system (i.e. the Fault Tree-FT and the Multi-States Physics-Based Model-MSPM) are presented; finally, in Sect. 2.4.3, the proposed approach is applied to evaluate the trustworthiness of the two models.

2.4.1. The system

The RHR system of a typical PWR reactor is taken as reference. The RHR is mainly used to remove the decay heat (residual power) from the reactor cooling system and fuel during and after the shutdown, as well as supplementing spent fuel pool cooling in the shutdown cooling mode for some types of reactors [4]. As illustrated in Figure 2.2, the main components of the RHR system are: pumps, heat exchangers, diaphragms, and valves. According to previous studies, it was found that 23% of RHR system failures are due to pumps failures, 58% are due to valves failures, while the rest of RHR system failures are due to other components' failures [54].

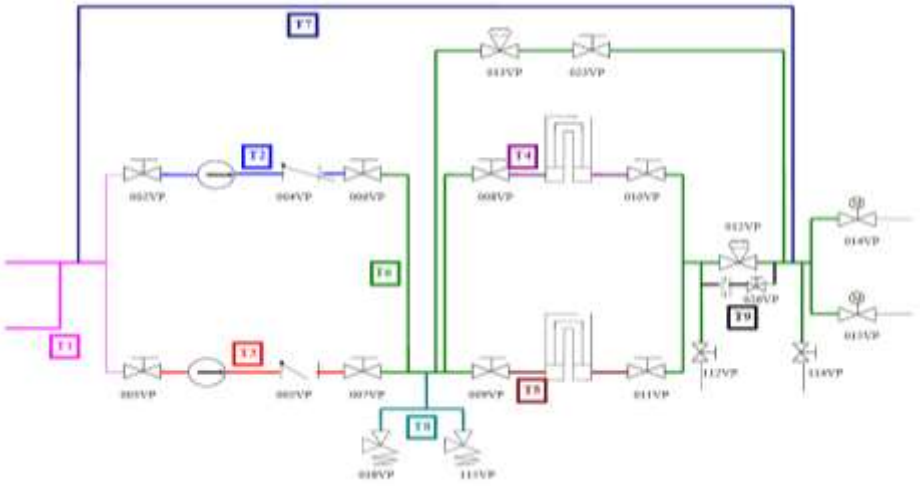


Figure 2.2 Schematic diagram of the RHR

2.4.2. Models considered

Two models have been considered for evaluating the reliability (resp., the failure probability) of the RHR system: a Fault Tree (FT) model (Sect. 2.4.2.1) and a Multi-State Physics-based Model (MSPM) (Sect. 2.4.2.2).

2.4.2.1. Fault Tree (FT) Model

Andromeda software has been used for the analysis of the RHR's components failure modes and criticalities (importance analysis). The analysis is based on a logical framework for understanding the different possible ways in which the components and the system can fail. The failure probabilities of the basic events used in the FT analysis are based on field experience feedback. The result of the FT analysis is

given in Table 2.3.

2.4.2.2. Multi-State Physics-based Model (MSPM)

The MSPM has been used for the analysis of the RHR's failure. In MSPM, the state transition rate estimates are based on Physical Based Models (PBM) rather than operational data [55], and the whole process of transition and degradation is, then, described by Multi-States Models (MSM) [56].

In the present analysis of the case study, the main critical components were taken into account (i.e. pump, diaphragm, breaker, motor, contactor and valve). The MSM was used to model the pump, breaker, motor and contactor, while the PBM model was used to model the valve and diaphragm, taking into account the degradation dependency of the valve on the pump.

More specifically, three states were considered for the pump, including the fully functioning state, a degradation state corresponding to external leakage and the failure state. The breaker was modeled by a continuous-time homogeneous Markov model, taking into account the perfectly functioning and the failed states, and four types of failures were taken into account. Similarly a continuous-time homogeneous Markov model was developed for the analysis of the contactor and the motor, and four and two types of failures were taken into account for each, respectively. On the other hand, the valve is subject to thermal fatigue that causes cracks or propagation of manufacturing defects, which are described by physical models and the related physical variables.

The results of MSPM and FT are given in Table 2.3. The analysis shows similarities results in the first eight years. A difference between the two results starts to appear in the tenth year, showing a more rapid decline in the reliability values obtained by MSPM.

Table 2.3 Values of reliability

Time (years)	0	1	2	3	4	5	6	7	8	9	10
Reliability/FT	1	0.779	0.607	0.473	0.369	0.288	0.224	0.175	0.143	0.107	0.083
Reliability/MSPM	1	0.775	0.603	0.469	0.366	0.285	0.222	0.173	0.135	0.105	0.060

2.4.3. Evaluation of model trustworthiness

The analysis is carried out through two main steps: the first is an "downward" evaluation of the weight of each element in the hierarchy tree with respect to the top goal of model trustworthiness; the second is a "upward" assessment of the model trustworthiness by means of a numerical evaluation of the basic "leaf" elements for both FT and MSPM models, as shown in Figure 2.3.

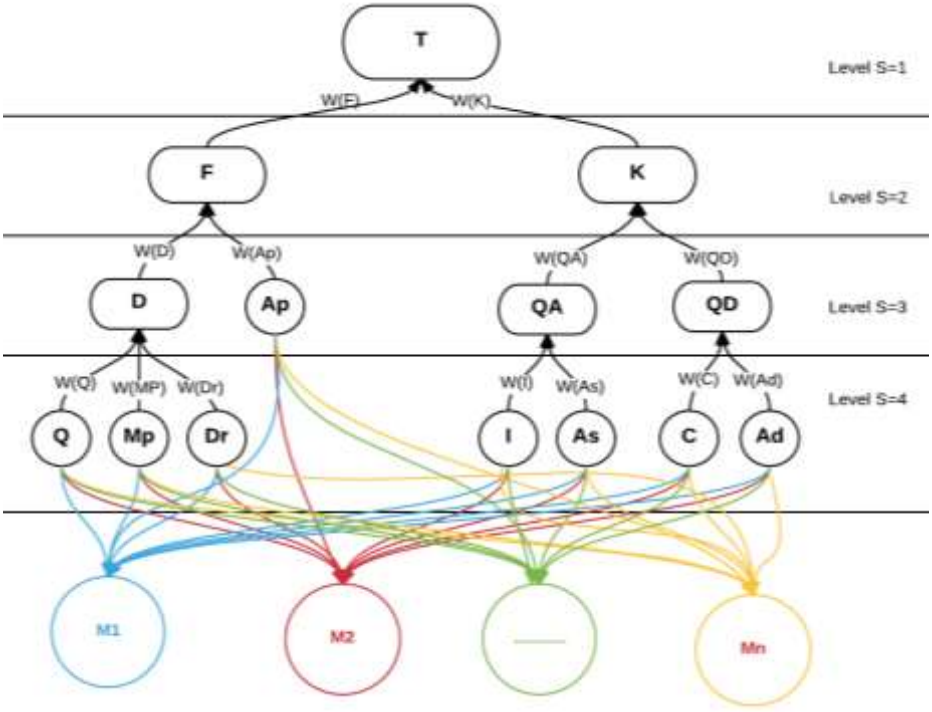


Figure 2.3 Hierarchical tree-based AHP model for the assessment of the trustworthiness of risk assessment models

With respect to the weights evaluation, three experts were asked to fill the pairwise comparison matrices, in order to evaluate the importance of each attribute (criteria). As the experts were considered equally qualified, the weights from different experts, were averaged. The results are presented in Table 2.4. In particular, the weights of each attribute with respect to the corresponding "upper level" parent (i.e., $W(T_i), W(T_{ij})$ *and* $W(T_{ijk})$) as well as the "global" weight $W_{global}(T_{ijk})$, with respect to top goal T are given. For more information on how to apply AHP method and solve the pairwise comparison matrices, please see Sect. 2.3.1 and the case study in the appended paper I.

The second step consists in an "upward" calculation, for the evaluation of the basic "leaf" attributes for each model. Actually, based on the data, information and knowledge available and used in the risk assessment analysis, two types of trustworthiness analysis have been implemented. In the first type, the analysis is performed through a direct quantitative evaluation of the leaf attributes (e.g., for **Mp** (T_{112}), the number of model parameters are counted, for each model if possible) or quantified semi-quantitatively if the attribute is qualitative in nature or doesn't correspond with the principle of the larger the better. In the second type, the analysis is based on a semi-quantitative evaluation of the leaf attributes carried out through comparing the two models to each other and to the state of the art, and then, assigning a relative score (1-9) for each leaf attribute.

In order to do that, scaling guidelines have been defined for the "leaf" attributes based on several EDFs' technical reports, [57] and the feedback of experts, and scores of 1-9 have been defined (see Appendix A of the appended paper I for details). Actually, we do not claim that those guidelines are complete and comprehensive, but they are sufficient for the context of the work. Relying on the guidelines of Appendix A of the appended paper I, the data and technical reports used to perform the risk assessment, the relative score evaluation was performed for both FT and MSPM models: the results are reported in Appendices of the appended paper I, respectively. In passing, notice that the evaluation of the attribute "Impact of the assumptions" ($I = \mathrm{T}_{212}$) is made as follows: a scale is given for each assumption and the scores are, then, averaged over all the assumptions.

On the basis of the relative scores selected, the trustworthiness evaluation was performed for both models, as illustrated in Table 2.4: the "normalized" level of trustworthiness was found to be 0.44 for Ft (M_1) and 0.56 for MSPM (M_2) by relative semi-quantitative evaluation of the attributes. Whereas they were found to be 0.34 for M_1 and 0.66 for M_2 by the quantitative evaluation.

We have applied the same method also to evaluate the models trustworthiness T using the direct quantification of the leaf attributes. The results are reported in Table 2.4.

Table 2.4 Comparison between FT and MSPM trustworthiness (relative/direct quantification)

Attribute	Weight	Global weight	Relative scores				quantitative evaluation			
			Fault Tree		MSPM		Fault Tree		MSPM	
			S	WS	S	WS	S	WS	S	WS
T	1.00	1.00	-	4.65	-	5.85	-	58.45	-	113.59
F (T_1)	0.35	0.35	-	1.51	-	2.37	-	1.67	-	2.66
Ap (T_{12})	0.54	0.19	6	1.13	7	1.32	7	1.32	7	1.32
D (T_{11})	0.46	0.16	-	0.38	-	1.04	-	0.35	-	1.34
Q (T_{111})	0.46	0.07	3	0.22	8	0.60	1	0.07	9	0.67
Mp (T_{112})	0.21	0.03	3	0.10	7	0.24	8	0.27	18	0.61
Dr (T_{113})	0.32	0.05	1	0.05	4	0.21	0	0.00	1	0.05
K (T_2)	0.65	0.65	-	3.14	-	3.49	-	56.78	-	110.93
QD (T_{22})	0.51	0.33	-	2.06	-	2.25	-	55.76	-	109.89
Ad (T_{221})	0.60	0.20	5	0.99	8	1.59	275	54.70	549.15	109.23
C (T_{222})	0.40	0.13	8	1.06	5	0.66	8	1.06	5	0.66
QA (T_{21})	0.49	0.32	-	1.08	-	1.23	-	1.02	-	1.04
As (T_{211})	0.20	0.06	5	0.32	6	0.38	4	0.25	3	0.19
I (T_{212})	0.80	0.25	3	0.76	3.33	0.85	3	0.76	3.33	0.85
*S: score		*WS: weighted score								

2.5. Conclusion

In this work, we have developed a hierarchical tree-based decision-making framework to assess the relative trustworthiness of risk models. The approach is based on the identification of specific attributes that are believed to affect the trustworthiness of the model. This is obtained through a hierarchical-tree based "decomposition" of the model trustworthiness into sub-attributes. The AHP method has been used to perform a weighted aggregation of the attributes to evaluate the model trustworthiness. The method has been applied to a case study involving the RHR system of a NPP. Two models of different complexity (i.e., FT and MSPM) have been considered to evaluate the system reliability and the trustworthiness of such models has been compared.

FT trustworthiness has been found to score 4.65 out of 9, whereas MSPM has scored 5.85 out of 9 by the relative semi-quantitative evaluation of leaf attributes (or 0.34 and 0.66, respectively, by normalizing the results). Please note that 9 the maximum score in the scaling system. The quantitative evaluation of the two models resulted in 58.45 for FT, whereas 113.59 for MSPM or 0.56 and 0.66 when normalized. The two results confirm the expectation that MSPM provides more trustworthy risk estimates than FT, due to the fact that it takes into account components failure dependency relations and time dependency of the degradation affecting the component.

Clearly, there is no claim that the trustworthiness assessment approach proposed is comprehensive and complete, as there exist other factors that affect the level of trustworthiness, which were not considered here. The method was, rather, a first attempt to systematically evaluate the models' relative trustworthiness. Obviously, it impossible to remove completely subjectivity and expert judgment is still present, the method provided is an attempt to cast such expert judgment in a systematic and structured framework. Also, further studies should be performed to define the scaling guidelines for attributes evaluation and study how to integrate the level of trustworthiness in RIDM.

Chapter 3 Risk analysis model maturity index for Multi-Hazards Risk Aggregation

In risk assessment, we measure risk quantitatively or qualitatively to inform design solutions and maintenance strategies so that the risk is maintained below the accepted limit. The evaluation of the overall risk implies aggregating the risk indexes from different contributors, i.e., MHRA.

MHRA must be capable of combining the outcomes of the risk assessment models relative to the different contributors, which are heterogeneous in nature and based on different degrees of maturity [19]. The current practice of MHRA adopts a simple arithmetic summation of the risk outcomes relative to the different contributors, without considering the different levels of knowledge base and maturity of the models used to obtain them [19]. The current practice of MHRA should be extended to reflect the level of maturity of the different risk analysis models whose outcomes are involved in the aggregation. In this chapter, a new index, namely the level of maturity, is introduced to reflect factors of heterogeneity in the assessment of the different risk contributors involved in the MHRA. A review of approaches for MHRA proposed in the literature is presented in Sect. 3.1. In Sect. 3.2 we propose a hierarchical tree to structure the level of maturity of a risk assessment model, we discuss the effect of the factors influencing the level of maturity on the risk assessment and DM and propose some evaluation guidelines. In Sect. 3.3, we illustrate how to evaluate the level of maturity for a given hazard group and introduce the reduced-order model to allow application on large scale PRA models. In Sect. 3.4, we apply the developed methods on a numerical case study. Finally, in Sect 3.5, we give conclusions and discuss potential future work.

3.1. State of the art

Few works in the literature focus on MHRA in risk assessment. EPRI report [19] indicates that the current practice of MHRA might not be appropriate for some contexts of DM due to the difference in the means employed for evaluating risk and the degrees of confidence in the risk contributors. The report also highlights some of the fundamental differences in the nature of the risk estimates from different sources (e.g., maturity of the used tool and analysis, uncertainty level for each contributor) [19]. Then, it proposes a practical guidance for an integrated understanding of the risk to support DM within the context of

RG1.174 (i.e., meets the current regulations, meet the defense in depth requirement etc. See Figure 2 in [33] for more information). This is done by developing the relevant insights for each of the contributions to risk. Five main tasks are required and "iteratively" performed in this guidance: (i) understand the role of PRA in supporting the decision; (ii) identify the main risk contributors and evaluate the baseline risk and assessing the credibility or confidence in the assessment; (iii) evaluate relevant risk metrics and refine the PRA if needed; (iv) identify and characterize key sources of uncertainty; (v) document conclusions for integrated DM. It should be noted that this work doesn't provide a clear guidance on evaluating the level of realism and trustworthiness in risk assessment.

An iterative method is also proposed in [34] for assessing the different aspects of risk aggregated from highly heterogeneous hazard groups and provide useful insights for RIDM, focusing on relative rather than absolute risk metrics. The method uses response surfaces that are based on arbitrary polynomial chaos expansion in combination with radar charts to visualize the overall risk and associated uncertainties. The response surface allows identifying major contributors to the overall risk, individually or on aggregate bases for a very large number of input parameters. On the other hand, radar charts are used to visualize risk contributors of different natures and compare them to safety guidelines. The method allows the comparison of risk contributors. However, it does not address factors like model conservatism, biases, incompleteness, hidden model uncertainty (e.g., structural), etc. Also, radar charts do not really allow the aggregation of risk from different contributors. Instead, they only allow the relative comparison of the risk contributors (hazard groups) to a given threshold.

3.2. A hierarchical framework for PRA maturity assessment

In risk assessment, many factors are believed to affect the suitability of risk definition and risk aggregation. Emphasis is paid in the literature on importance of communicating these factor for better informing DM [6], [36], [17], [19], [41]. In particular, MHRA includes aggregating risk indexes from different contributor that have different degrees of realism [19]. Different aspects leading to heterogeneity in the realism of risk analysis are identified in the literature. Some of these aspects are: (i) background knowledge; (ii) level of uncertainty; (iii) level of conservatism; (iv) importance measures; (v) level of details and sophistication of the analysis; (vi) accuracy and precision in the estimation of the values of the model's parameters; (vii) level of sensitivity; (viii), and others [1], [6], [36], [8], [17], [3], [19], [41], [58], [11].

In this section we propose a conceptual hierarchical tree to evaluate the maturity index based on some

attributes that are believed to affect the level of maturity of the risk analysis and that have huge implication of DM (Sect. 3.2.1). In Sect 3.2.2 we demonstrate the implication of these attributes on the maturity and propose scoring protocols for the evaluation of the attributes.

3.2.1. The developed framework

In this work, we focus on communicating the factors that lead to heterogeneity in the estimation of the different risk indexes, and accordingly affect their degrees of realism, through a metric referred to as "level of maturity". The level Maturity of a PRA expresses the degree to which PRA is correctly implemented in a way that makes best use of the available knowledge to best represent the reality.

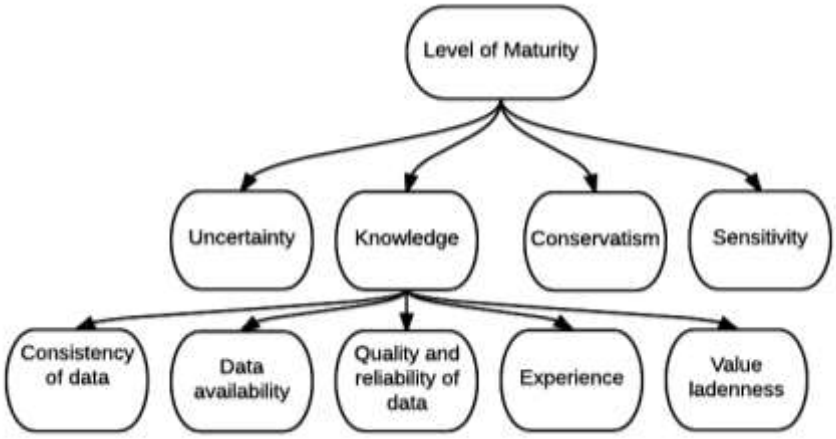


Figure 3.1 Level of maturity framework

In this work, four elements i.e., uncertainty, conservatism, knowledge and sensitivity [1], [6], [36], [19], [58], [11] relevant to the level of maturity and RIDM are reviewed and discussed. In this discussion, we argue the importance of these attributes in determining the level of realism of probabilistic risk analysis and we propose evaluation protocols that are based on solid argument presented in the same sections. The overall hierarchical representation of the framework is illustrated in Figure 3.1.

3.2.2. Attributes evaluation

In this section, we review the elements presented in Figure 3.1 and discuss their implication on the maturity of risk assessment and accordingly propose evaluation procedures.

3.2.2.1. Uncertainty

Uncertainty is defined as the imperfection of knowledge on the real value of a variable or its variability [59]. Uncertainty is an important source of differences between the reality and the model predictions [3]. Hence, uncertainty affects greatly the credibility of PRA [60], [61]. This means that it

reflects directly the level of maturity of the PRA and it should be addressed in its evaluation.

Uncertainty classification

Uncertainty can be classified relatively into different levels, depending on the degree of knowledge imperfection [62]. For example, [63] distinguishes four types of uncertainties depending on the level of knowledge: *"Risk"* where the system behavior is well known and quantifiable; *"uncertainty"* where the system parameters are known but the probability distributions are unknown; *"ignorance"* where the unknowns are unknown and finally; *"indeterminacy"* which underlies the indeterminacy in scientific knowledge. Walker et *al.,* (2003) suggests three dimensions for uncertainty classification for uncertainty-based decision support purposes: the *"location"* where the uncertainty manifests itself within the model complexity, the *"level"* of uncertainty, which is, demonstrated by a spectrum between deterministic knowledge and absolute ignorance and finally, the *"nature"* of uncertainty which illustrates the type of uncertainty (epistemic or aleatory) [62]. The level of uncertainty is, further, classified into five progressive levels: determinism, statistical uncertainty, scenario uncertainty, recognized ignorance and total ignorance [62]. Spiegelhalter and Riesch (2011) identify, within the spirit of [62], five progressive levels of uncertainty for model-based risk analysis [64]. The levels are presented in Table 3.1.

Table 3.1 Uncertainty levels descriptions and scores with respect to the level of maturity

Level	Description	Score
Level 1 (uncertainty about the outcome)	This level of uncertainty manifests itself when the model and the parameters are known, and the analysis predicts a certain outcome with a probability P (e.g., the uncertainty about the outcome in most traditional mathematical and philosophical problems of probability theory)	5
Level 2 (uncertainty about the parameters	The model is known but its parameters are not. If the parameters are known then the model would predict an outcome with probability P and exhibit an uncertainty of level one. This type of uncertainty arises due to lack of empirical information on the model parameters (e.g. input parameters related to Large Break in Primary Circuit of a Nuclear Power Plant that has never occurred)	4
Level 3 (uncertainty about the model)	It reflects the likelihood of the competing models' abilities to reflect reality. This type of uncertainty is due to the model structure itself and the computer implementation of the model [62]	3
Level 4 (uncertainty about the acknowledged limitations and implicit assumptions-unmodeled	This level covers any known limitations in understanding and modeling abilities, which arises from the inevitable assumptions and simplifications made such as: data extrapolations, limitation in the computations, and any aspects that we are aware that they have been omitted.	2

uncertainty)		
Level 5 (Uncertainty about unknown inadequacies)	It is the unrecognized uncertainty or as it was referred to by Donald Rumsfeld the "unknown unknowns", which corresponds to the unforeseen events, unmodeled and unmodlable uncertainty. This type of uncertainty are usually acknowledged by brainstorming of the possible scenarios, or by the introduction of what so-called 'fudge factors'.	1

Whilst this classification seems crude and simple, it satisfactorily covers, at least from this problems' perspectives, the three dimensions defined by [62], i.e., *"location"*, *"level" and "nature"* of uncertainty. For example, the definition of uncertainty Level 1 refers to the aleatoric nature of uncertainty, whereas Levels 2-5 cover the epistemic nature of uncertainty. Also, where the five levels vary progressively from the known to the unknown-unknown, they simultaneously refer to its location i.e., parameter, model and context of uncertainty. Please notice that classification can be applied on the level of the hazard group as well as on the level of the basic events in the PRA model since the probabilities of basic events are determined using data and physical or statistical models.

3.2.2.2. Conservatism of analysis

Conservatism in PRA refers to desire of overestimating the risk purposely out of cautiousness. The conservatism in PRA arises from different considerations and perspectives, such as the concerns regarding the lack of knowledge about the nature and magnitude of the hazard [65]. This leads to the implementation of the concept of "Better safe than sorry", which is further translated to the preference of overestimating the risk rather than underestimating it. For example, selecting risk estimate at the 95^{th} percentile, which, means that there is a 95% probability that the risk is overestimated and 5% is underestimated [66].

Although the conservatism is usually anticipated to increase safety, some counter-arguments still exist on its influence on safety margin [66]. It has been argued that conservatism cannot be advised only from a risk-aversion point of view, and that the cumulative effects of conservatism on decision-making, regulations and risk management are unacceptable [66], [65]. In particular, the effect of conservatism is not taken into account from a firm empirical sense [65], which might be, in some contexts, perceptive for the analysts by giving a false assurance of safety, leading to worst consequences of risk [67]. In fact, the overall effect of conservatism on safety (whether that conservatism is protective or not), depends greatly on the assumptions made, and the context of DM [67].

Viscusi *et al.* (1997) argue that though conservative risk estimates increases the risk magnitude, the implications of this increase on the safety is still a matter of the decision-makers' actions [65]. They have showed through a cost-benefit-based study (number of lives saved per unit cost) that unlike conservative

assessment, the mean parameter approach would result in enhanced judgment policies that would enhance the safety. This can be explained by the shift of prioritization of decision maker. Moreover, recent studies conclude and explicitly recommend that conservatism should be avoided in the light of some DM contexts like: comparing options and studying the effects of potential risk reduction measures [58]. The degree of conservatism should be complied with the decision contexts and requirements of the PRA. Otherwise, it might reduce the maturity level and sometimes mislead the decision maker.

Conservatism classification

All of the arguments mentioned in the previous section lead to questioning how to classify of levels of conservatism in the light of the maturity and its consequences on safety. At a first glance, classifying the levels of conservatism depending on the level of knowledge seems plausible, especially that conservatism represents a practical act performed to deal with uncertainties and lack of knowledge. However, this is not valid considering its implication on safety, where other aspects should be taken into account aside from strength of knowledge, e.g., the context of DM. Aven (2016) highlights the conservatism in risk analysis as a multi-dimensional concept, reinforcing the former arguments of experts about the real effect on safety [58]. This is done by firstly addressing the meaning of conservatism, secondly relating it to the strength of knowledge and thirdly evaluating its usefulness in the context of decision-making. In this vision, he compares conservative risk indexes (i.e., based on conservative assumptions) to three cases: (i) risk indexes based on best estimate assumptions; (ii) risk indexes based on true value parameters (iii) risk indexes based on true value parameters with a defined confidence statement. Then, for these cases (i-iii), he defines the possible states of knowledge on which the assumptions or risk parameters are based and finally, the possible contexts of decision, and tries to relate it to the consequences on safety [58]. Hereafter, we extend the work in [58] and define three main types of risk index estimates: (i) best judgment estimates (based on best judgment of assumptions and parameters); (ii) true value with a high confidence (based on strong knowledge); (iii) true value with a low confidence (based on weak knowledge). Then, for two context of DM, i.e., comparing alternatives and comparing the risk indexes to acceptance limit, we compare the three defined estimate types (i-iii) to the conservative estimates (based on conservative assumptions) and give scores for each possible scenario with respect to level of maturity and safety. In other words, we are comparing the estimates that are based on assumptions chosen to be conservative (for cautiousness reasons) to those estimates that are based on the best judgment or true values of assumptions and parameters. Figures 3.2-3.4 illustrate the different score for each corresponding scenario. From Figures 3.2-3.4, five levels of conservatism are defined in light

of their influence on the safety, where Leve l represents the worst influence of conservatism in terms of reducing the safety, Level 3 represents an acceptable influence of conservatism on safety, Level 5 represents the best influence of conservatism on increasing the safety. Levels 2 and 4 are intermediate levels.

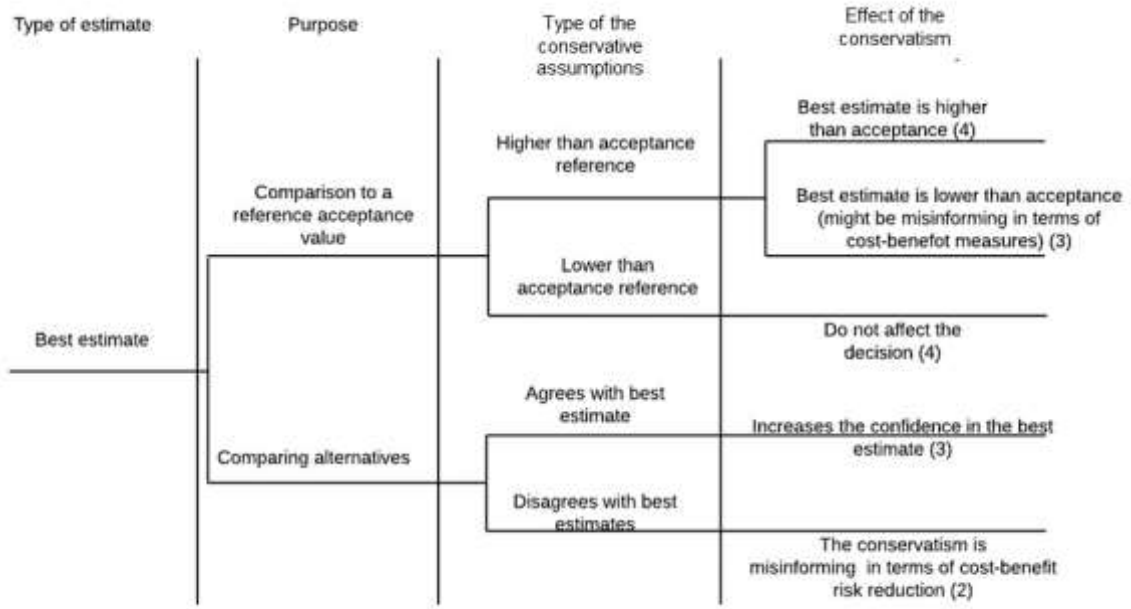


Figure 3.2 Evaluation of the conservatism in the light of level of maturity (conservatism VS Best estimate)

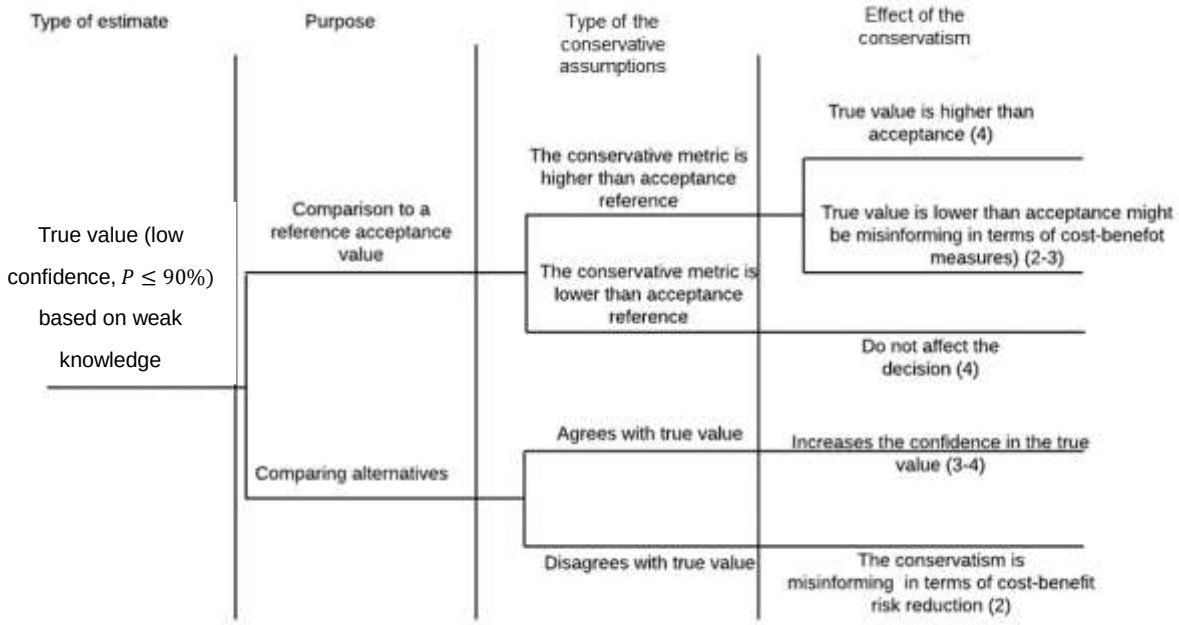


Figure 3.3 Evaluation of the conservatism in the light of level of maturity (conservatism VS True value/weak

knowledge)

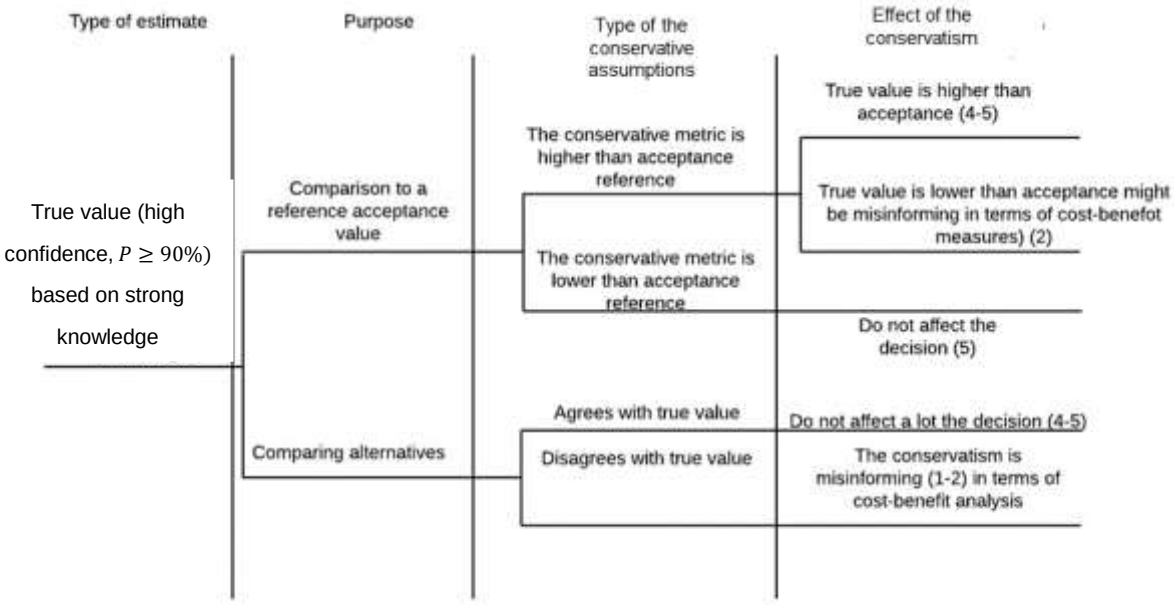


Figure 3.4 Evaluation of the conservatism in the light of level of maturity (conservatism VS True value/strong knowledge)

3.2.2.3. Knowledge

Knowledge is the second top tier of the four levels knowledge-hierarchy (DIKW hierarchy). It is the yield of a combination of data, information, experience and judgment to be used in decision-making [8]. Knowledge manifests itself in three main forms: explicit, implicit, and tacit [68].

It is said that “You can't manage what you can't measure". To best employ knowledge, one should be able to state its level. This led experts in safety and risk assessment to emphasize the importance of considering the background knowledge on which risk assessment is based, especially for RIDM purposes [8], [17], [18], [11], (Askeland et al., 2017), [16], [26]. This argument is visibly manifested in the new risk perspectives, which considers strength of knowledge in addition to the traditional elements i.e., scenarios, likelihood and consequences [17], [18], [69], [70]. For these reasons, evaluating strength of knowledge should be considered in evaluating the models’ credibility and maturity.

Knowledge evaluation

Different attributes can be considered to evaluate the strength of knowledge, such as the amount of data and information, its suitability and usefulness, the human cognition regarding a specific phenomenon, the experience on the technology and of the analysts, etc. There are, however, two main methods on which most of the strength of knowledge assessment approaches are based: a semi-quantitative approach for evaluating

the strength of knowledge [43] and the assumption deviation risk by [17]. In the first method, four main criteria are identified for evaluating the strength of knowledge: the phenomenological understanding, the reasonability and realism of assumptions, the availability of reliable and relevant data and the agreement among peers [43]. Based on the degree of fulfilling the criteria, the strength of knowledge is classified crudely to minor, moderate, and significant. The second method is based mainly on evaluating the criticality of the main assumptions on which probabilistic risk assessment is based. This is done by evaluating three criteria: deviation from assumption, the uncertainty of this deviation and the strength of knowledge supporting the assumptions. Accordingly, the number of assumptions and the criticality of deviation from assumption, indicates the strength of knowledge on which the probabilistic risk assessment is based [17]. However, one should not forget that in addition to the explicit properties of knowledge, it has also implicit and tacit properties [68]. Although it cannot be directly stated or documented, it contributes to the individual and organizational performance [71]. Obviously, in [6], the reasonability of assumptions and agreement among peers are partially related to the implicit and tacit knowledge. However, this framework does not cover convincingly the assessment of tacit knowledge (e.g., agreeing on an assumption or assessment does not necessarily make it good). Hence, the carriers of implicit and tacit knowledge (assessors) should rather be themselves evaluated.

In fact, several researches have emphasized on the importance of evaluating the value-ladenness and confidence in experts' judgment. For example, [24] points to the fact that expert's judgment is subject to inevitable bias that lead experts that have the same background knowledge to make different judgment. It defines a few attributes that are believed to affect the experts' judgment, such as, the personal interest, the personal knowledge, the degree of independence, the experience, etc. Other aspects such the situational limitations, choice space, agreement among peers and stakeholders are included as well to assess the quality and robustness of assumptions on which risk analysis is based [53], [21], [10]. Above all, one can argue that there are many other attributes that could be used to better represent the level of knowledge.

The method discussed earlier, which relies on four criteria for evaluating the strength of knowledge (i.e., the phenomenological understanding, the reasonability and realism of assumptions, the availability of reliable and relevant data and the agreement among peers [43])) seems very plausible and relevant to the context of this problem except that it doesn't take into account the assessment of the experts who make the assumptions and the reasoning of the analysis, neither the availability of trustable predicting models. In this work, we adjust and expand this method in Table 3.2, and add a new main attribute i.e., value-ladenness

of the assessor to the framework, to be adapt to the context of this chapter.

Table 3.2 Level of knowledges' attributes evaluation guidelines

	Score	**1**	**3**	**5**
Data availability *(A)*	Amount of data/field data ($Sc_{3,1}$)	No data or the data are so limited and (can extracted only from the same type of NPPs)	The data are available and can be extracted from any other NPP	The data are available in abundance (can be extracted easily from so many sources and places worldwide)
Data consistency *(Co)*	Source of data ($Sc_{3,2}$)	The data are extracted from other sources that is not related directly to the technology (not the exact same type of component)	Other NPPs of the same type and technology	Field data from the same power plant, and related to the same type of components
Quality and reliability of data *(Q)*	Quality of data ($Sc_{3,3}$)	Based on experts elicitation / Data are calculated using statistical models	Data are both assumed and calculated using computer physical and mathematical models	Data are extracted using computer mathematical and physical models / The data are measured precisely and accurately, and then modeled
	Quality of assumptions ($Sc_{3,4}$)	Represents strong simplifications	Represents moderate simplifications	Represents reasonable simplifications
Experience *(E)*	Phenomenological understanding ($Sc_{3,5}$)	The phenomena involved are not well understood	The phenomena involved are understood but not completely	The phenomena involved are very well understood
	Experience and knowledge regarding the hazard group ($Sc_{3,6}$)	No experience at all	Experienced such an event in other industries	This event is quite common and we have a wide experience in
	Availability of models ($Sc_{3,7}$)	Models are non-existent or known to give poor predictions.	The models used are believed to give predictions with moderate accuracy	The models used are known to give predictions with the required accuracy
Value ladenness of the analysts *(VL)*	Agreement among peers ($Sc_{3,8}$)	There is strong disagreement among experts	There is slight agreement among experts	There is broad agreement among experts
	Expert years in experience in the field and performance measure ($Sc_{3,9}$)	has quite short experience in risk assessment of NPPs	It is his specialty and he practiced through training courses regarding the same type of NPPs	Expert in this domain (long experience)

3.2.2.4. Sensitivity

A mathematical model might embrace errors due to the lack of the knowledge regarding the input parameters or due the numerical methods used to solve the model [72]. The effects held by such errors are very important and need to be evaluated as it reflects the range of the trustworthiness and validity of the model. This is, done by sensitivity analysis [72].

Sensitivity analysis is generally used to determine how a dependent variable can be changed and affected by the change of the input independent variable [72]. This is usually used to determine the critical control points and to prioritize additional data collection [73]. Moreover, it is implemented to provide the comprehensive understanding needed for a reliable use of the model, through highlighting and quantifying its most important features [72], as well as verifying and validating it [73].

In safety and risk assessment, sensitivity analysis can be useful in many ways. In particular, sensitivity analysis complements the risk analysis to inform decision-making [74], where it helps to identify the uncertain inputs that contributes to the uncertainty in the outputs and consequently, affect the DM process [75]. For example, in PRA of NPPs, sensitivity analysis is required to study the impact of different model basic events' probabilities on the decision [76]. Also, the importance of an assumption in a risk prediction model can be evaluated through altering the input parameters or the background knowledge related to the given assumption, which helps in identifying the critical assumptions and the risk of their deviations [43]. Furthermore, sensitivity analysis is recommended in the practice of risk assessment to reduce -in some cases- the unnecessary conservatism [33]. From these perspectives, sensitivity analysis is considered an indispensable tool for evaluating model credibility and maturity.

Sensitivity evaluation

Flage and Aven (2009) suggested integrating the sensitivity concept as a main component of the uncertainty in order to have a holistic picture of the uncertainty beyond the concept of the probability [6]. A rough semi-quantitative evaluation of sensitivity has been introduced with three levels of classification: significant sensitivity, moderate sensitivity and minor sensitivity. The simplicity of this method makes it very helpful in the context of DM, as it gives an indication on the associated consequences and implications of parameters' deviations. On the other hand, it doesn't show how to apply the sensitivity analysis, neither how to translate it into a sensitivity level. For this reasons, we suggest to complement this proposal by using a one-at-a-time index and then, converting it into a relative scores that represents the sensitivity levels.

In one-at-a-time method, the sensitivity index S measures the relative change in the dependent (output) variable $Y(x_i)$ by altering one input (x):

$$S = \left|\frac{Y(x_1+\Delta)-Y(x_1)}{Y(x_1)}\right|, \tag{3.1}$$

where x_i is the input parameter, Δ is an estimated suitable value by which the input parameter is alerted e.g., ±20% of the original value, ±SD (standard deviation) [77] or ±4SD [78]. However, we are considering a ±50% altering parameter in this study to represent more clearly the sensitivity of parameters, as we are more concerned with PSA models that have a linear relation with the basic events (if each basic event is unique and appears only one time in a given minimal cutset).

In this kind of analysis converging to 0 indicates the insensitivity of the model, while diverging from 0 indicates sensitivity. After applying these analysis, the results need to be converted into discrete scores (e.g., 1: minor, 2: moderate, 3: significant [43]) that indicate their levels. A sensitivity score (1-5) is assigned for the sensitivity index relying on the degree that the index converge or diverge from 0 as illustrated in Table 3.3. Please note that mapping the sensitivity indexes into scores is based on subjective elicitation and can be adapted given the context.

Table 3.3 Scores representation of the sensitivity measure

Interval	S: ≤0.10	S : 0.10-0.25	S: 0.25-0.45	S: 0.45-0.70	S: ≥0.70
Level of sensitivity	1	2	3	4	5
Score	5	4	3	2	1

Please notice that if we are applying the sensitivity analysis on the level of the basic events of the PRA model, then, it means that we are studying the dependency of the PRA model on this given basic event.

3.3. PRA maturity assessment

In this section we implement the developed framework through Analytical Hierarchy Process (AHP) method in Sect. 3.3.1. Then, we develop a method for evaluating the level maturity on the basis of small constituting elements of the PRA model in Sect. 3.3.2-3.3.4. In Sect. 3.3.5 develop a technique for aggregating the maturity of the overall risk analysis.

3.3.1. Evaluation of the level of maturity

For each criterion and sub-criterion defined in Figure 3.1, a semi-quantitative evaluation is carried out by assigning a relative score from 1 to 5, based on the set of pre-defined scoring criteria presented earlier in Sect. 3.2.2. The next step is to aggregate the scores of different attributes (criteria) to assess the overall

maturity of a risk contributor. In this work, the maturity level is calculated as a weighted average of the scores of the attributes.

$$m_i = \sum_{j=1}^{N_p} \sum_{i=1}^{n_d} w_i \cdot w_{i,j} \cdot Sc_{i,j} \tag{3.2}$$

where m_i is the level of maturity for the i-th hazard group that need to be evaluated, $w_{i,j}$, Sc_{ij} and w_i are respectively the weight and the score of the j-th sub-attribute in the i-th attribute, and the weight of the i-th attribute. N_p is the total number of attributes and n_d *is* the number of sub-attributes related to the i-th evaluation criterion. The relative weight of each attribute w_i and sub-attribute $w_{i,j}$ is determined by Analytical heretical Process (AHP). Detailed description of AHP method was introduced in Chapter 2.

3.3.2. The concept of reduced order model

After determining the relative weight of the attributes, Eq. (3.2) can be applied to determine the level of maturity. Evaluating the level of maturity on the level of hazard group, however, is not realistic. Further, PRAs of complex systems are very complex and often embrace multiple PRA elements, which need to be evaluated separately. In this light, we develop a technique to limit the number of elements that need to be analyzed PRA models, namely, the reduced-order model.

For the purpose of illustration, we consider the Probabilistic Risk Assessment (PRA) models used in the nuclear industry. Specifically, we refer to the widely applied event tree models. The events probabilities in the event tree model are calculated by fault tree models. The risk index considered is the probability of occurrence of a given consequence (e.g. the probability of core damage in a NPP). For each combination of operation state and scenario, a dedicated risk assessment model (in this case, an event tree) is developed and the total risk index is calculated by summing the values of the risk indexes calculated for each individual risk model:

$$R = \sum_{i=1}^{n_O} \sum_{j=1}^{n_{S,i}} R_{i,j}, \tag{3.3}$$

where n_O is the number of operation states (O), $n_{S,i}$ is the number of accident sequences (scenarios, S) that are considered in operation state i and can lead to the given consequence of interest. Each $R_{i,j}$ in Eq. (3.3) quantifies the risk contribution specific to scenario j (e.g., medium flood level) in operation state i (e.g., emergency shutdown).

The risk models for calculating the specific risk index contribution $R_{i,j}$ are characterized by initiating events (IEs), basic events (BEs) and their combinations in minimal cut sets (MCSs). Please note that the initiating events in the PRA model are basic events that trigger the abnormal activity, so it will be

treated hereafter as a basic event. Taking the rare-event approximation, $R_{i,j}$ can be calculated by [79]:

$$R_{i,j} = \sum_{k=1}^{n_{MCS,i,j}} \prod_{q \in MCS_k} P_{BE,q}, \tag{3.4}$$

where $n_{MCS,i,j}$ is the number of minimal cut sets in the risk model for operation state i and scenario j, MCS_k is the k-th minimal cutset and $P_{BE,q}$ is the occurrence probability of the q-th basic event in MCS_k.

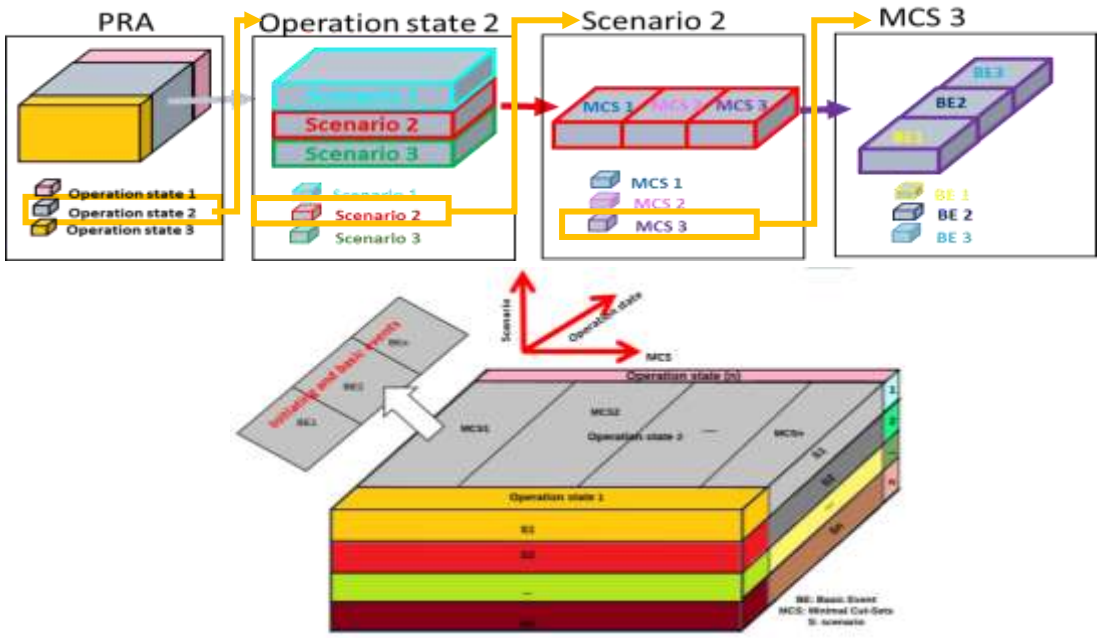


Figure 3.5 Atomic elements of a PRA model

For the following illustration of the maturity assessment procedure, it can be considered that the four elements O, S, MCS and BE fully define the PRA model, as shown in Figure 3.5. We refer to these four elements as the "constituting elements" of the model. In Figure 3.5, let's imagine that the PRA model is a box (cuboid). The box is divided into several cuboids each represents a given operation state. Each operation state cuboid is further broken down into smaller cuboids that represent the scenarios. The scenario cuboids are in turn broken into smaller cuboids each represents a MCS. Finally, the MCS cuboids are broken into the smallest constituting cuboids (known as the basic atomic elements) that represent the basic events. The idea behind this technique is to facilitate the process of maturity evaluation by dividing the PRA model into the smallest constituting elements known as the atomic elements. As illustrated in Figure 3.5, the atomic elements of the PRA model are the basic events.

To assess the maturity of the PRA model, all the four atomic elements must be considered. In practice, however, PRA models are very complex: they contain many scenarios and operation states, combined in

large and complex fault trees and event trees, that consist of thousands of BEs and MCSs [80]. For such complex risk assessment models, it is not practical to consider all atomic elements for evaluating the maturity. To address this problem, we develop a top-down bottom-up method for maturity assessment. A reduced-order model for Eq. (3.4) is developed first, in order to limit the number of atomic elements that need to be analyzed. The model allows the assessment of maturity for most basic atomic elements and then calculating it for the other constituting elements. A detailed discussion on how to construct the reduced-order model is given in Sect. 3.3.3. Then, the maturity supporting each atomic element in the reduced-order model is assessed by a weighted average of the scores for the attributes in Figure 3.1. The weights are evaluated using pairwise comparison matrices of AHP. In Sect. 3.3.4, the maturity of each element is aggregated to evaluate the maturity of the entire PRA model. Finally, an approach is presented in Sect. 3.3.5 for risk aggregation considering the level of trustworthiness.

3.3.3. Reduced-order PRA model construction

In PRA models, most of the contribution to the total risk is provided by a small number of basic elements (known as "*Pareto principle*") [81]. The rest of the basic elements might be in large number but contribute little to the total risk. To make feasible the maturity assessment, the PRA model is transformed into a reduced-order model that consists of the most important "*atomic elements*", in order to reduce the number of elements that need to be analyzed.

The procedure for constructing the reduced-order model is made of three steps. Firstly, the number of operation states n_O is reduced to the $n_{O,Red}$ most relevant; to do this:

- Calculate the risk R_{O_i} for each operation state:

$$R_{O_i} = \sum_{j=1}^{n_{S,i}} R_{i,j}\,, \quad 1 \le i \le n_O, \tag{3.5}$$

where $R_{i,j}$ is calculated by Eq. (3.4).

- Rank R_{O_i} $1 \le i \le n_O$ in descending order.
- Find the minimal $n_{O,Red}$, so that:

$$\frac{\sum_{i=1}^{n_{O,Red}} R_{O_i}}{R} \ge \alpha, \tag{3.6}$$

where α is the fraction of total risk that is represented by the operation states kept in the reduced-order model (in the case study in Sect. 3.4, we choose $\alpha = 0.8$).

- Keep only the first, most contributing operation states, i.e., those with $i = 1, \cdots, n_{O,Red}$; operation states with $i > n_{O,Red}$ are eliminated.

The second step is to define the reduced number of scenarios $n_{S,Red,i}$ for each operating state i in the reduced-order model, where $i = 1, \cdots, n_{O,Red}$:

- Calculate the risk $R_{i,j}$, $1 \le j \le n_{S,i}$ by Eq. (3.4).
- Rank $R_{i,j}$ in descending order, $1 \le j \le n_{S,i}$.
- Find the minimal $n_{S,Red,i}$ so that:

$$\frac{\sum_{j=1}^{n_{S,Red,i}} R_{i,j}}{R_{O,i}} \ge \beta, \tag{3.7}$$

where R_{O_i} is calculated by Eq. (3.5) and β is the fraction of total risk provided by the scenarios in the reduced-order model (in the case study in Sect. 3.4, we choose $\beta = 0.8$).

- Keep only scenarios for $j = 1, \cdots, n_{S,Red,i}$; scenarios with $j > n_{S,Red,i}$ are eliminated.
- Repeat the procedures for $i = 1,2, \ldots, n_{O,Red}$.

Finally, the number of minimal cut sets $n_{MCS,i,j}$ is tailored to $n_{MCS,Red,i,j}$, $i = 1, \cdots, n_{O,Red}, j = 1, \cdots, n_{S,Red,i}$:

- Calculate $R_{i,j,k}$ by:

$$R_{i,j,k} = \prod_{q \in MCS_{i,j,k}} P_{BE,q}\,, \quad {\scriptstyle 1 \le i \le n_{O,Red} \atop 1 \le j \le n_{S,Red,i} \atop 1 \le k \le n_{MCS,i,j}}, \tag{3.8}$$

- Rank $R_{i,j,k}$ in descending order.
- Find the minimal $n_{MCS,Red,i,j}$ so that:

$$\frac{\sum_{k=1}^{n_{MCS,Red,i,j}} R_{i,j,k}}{R_{i,j}} \ge \gamma, \tag{3.9}$$

where $R_{i,j,k}$ is calculated by Eq. (3.8) and γ is the fraction of total risk given by the minimal cutsets contained in the reduced-order model (in the case study in Sect. 3.4, we choose $\gamma = 0.8$).

- Keep only minimal cut sets for $k = 1, \cdots, n_{MCS,Red,i,j}$; minimal cut sets with $k > n_{MCS,Red,i,j}$ are eliminated.

Taking the rare-event approximation, the total risk of the reduced-order PRA model can be calculated by:

$$R_{Red} = \sum_{i=1}^{n_{O,Red}} \sum_{j=1}^{n_{S,Red,i}} \sum_{k=1}^{n_{MCS,Red,i,j}} \prod_{q \in MCS_{i,j,k}} P_{BE,q}, \tag{3.10}$$

Only the events that are contained in the reduced-order model (3.8) are considered when assessing the maturity. Note that from Eqs. (3.6), (3.7) and (3.9), the reduced order risk R_{Red} accounts for a portion $\alpha \times \beta \times \gamma$ of the total risk R. Please note that a value of 0.8 is usually chosen for α, β and γ (Pareto Principle). However, the assessor is free to adjust these values given the context of the problem.

From Eq. (3.10), the risk index of the reduced-order PRA model can be viewed as the sum of $n_l = \sum_{i=1}^{n_{O,Red}} n_{S,Red,i}$ risk index values $R_{Red,l}, l = 1, \cdots, n_l$ where $R_{Red,l}$ is known as the "elementary risk model" and calculated by the corresponding individual risk model, composed of MCSs and BEs at a given operation state and a given scenario, as shown in Eq. (3.11):

$$R_{Red,l} = \sum_{k=1}^{n_{MCS,Red,l}} \prod_{q \in MCS_{l,k}} P_{BE,q}, \tag{3.11}$$

In Eq. (3.11), $R_{Red,l}$ is the risk index of the l-th "elementary reduced-order risk model", where $n_{MCS,Red,l}$ is the number of MCSs in the l-th individual reduced-order risk model. In other words, the "individual reduced-order risk model" represents herby the risk model at a given operation state and a given scenario.

Assuming that the risk on reduced-order model is expressed by elementary reduced-order models, which represent the risk for each scenario at a given operation state, the weight of each elementary risk model can be expressed by:

$$W_l = \frac{R_l}{\sum_{l=1}^{n_l} R_l} \tag{3.12}$$

where R_l is the risk of elementary reduced-order model and n_l is the number of elementary reduced-order models and expressed by $n_l = n_O \times n_S$.

- Calculate the weight $W_{l,q}$ of each basic event in a given elementary reduced-order model by:

$$W_{l,q} = \frac{I_{l,q}}{\sum_{q=1}^{n_{l,q}} I_{l,q}} \tag{3.13}$$

where $n_{l,q}$ is the number of basic events in the l-th elementary reduced-order model, $I_{l,q}$ is the Fussell-Vesely importance measures of the q-th basic event in the l-th elementary reduced-order model.

3.3.4. Evaluation of the level of maturity of a single hazard group

Given the reduced-order model technique introduced in the previous section, the level of maturity can simply be evaluated by two steps:

- Evaluate the maturity on each basic event by:

$$m_{l,q} = \sum_{i=1}^{N_p} \sum_{j=1}^{n_d} w_i \cdot w_{i,j} \cdot Sc_{i,j,l,q} \tag{3.14}$$

where $m_{l,q}$ is the level of maturity for the q-th basic event in the l-th elementary reduced-order model, $w_{i,j}$ and $Sc_{i,j,l,q}$ are respectively the weight and the score of the j-th sub-criterion in the i-th evaluation criteria for the q-th basic event in the l-th elementary reduced-order model.

- Evaluate the maturity m_i for the total hazard group by:

$$m_i = \sum_{l=1}^{n_l} \sum_{q=1}^{n_{l,q}} W_l \cdot W_{l,q} \cdot m_{l,q} \quad (3.15)$$

3.3.5. Risk aggregation considering maturity levels

In this work, we adopt the perspectives of [17] that when characterizing risk, not only the probability index estimated by PRA, but also the knowledge that supports the PRA should be taken into account. Hence, in this work, we use a tuple (R_i, m_i) to quantify the risk associated with hazard group i, where R_i and m_i are respectively the risk index and is the maturity level of the i-th hazard group PRA model, evaluated based on the method presented in Sect. 3.3.1-3.3.4.

A two-stage aggregation method is, then, developed for MHRA considering maturities of hazard groups. Suppose we have n_h hazard groups with the risk tuple $(R_i, m_i), i = 1,2,\cdots,n_h$. The overall risk can, then, be represented as a risk tuple (R, M) and computed in two steps:

Step 1: Aggregation of risk indexes. Risk indexes are aggregated following the summation rule:

$$R = \sum_{i=1}^{n_h} R_i \quad (3.16)$$

where R is the risk index after considering all the hazard groups.

Step 2: Determine the maturity of the aggregated risk assessment:

In this work, the maturity can be represented for the overall framework by applying a weighted average the maturities from each hazard group, considering the risk contribution for each hazard group:

$$M = \sum_{i=1}^{n_h} W_{h,i} \cdot m_i = \sum_{i=1}^{n_h} \sum_{l=1}^{n_l} \sum_{q=1}^{n_{l,q}} W_{h,i} \cdot W_l \cdot W_{l,q} \cdot m_{l,q} \quad (3.17)$$

where $W_{h,i}$ is weight of the hazard group h and calculated as the following: n_h is the number of hazard groups in the risk assessment model:

$$W_{h,i} = \frac{R_i}{\sum_{i=1}^{n_h} R_i} \quad (3.18)$$

3.4. Application

In this section, we apply the developed framework on a case study of two hazard groups in NPPs. The level of maturity assessment framework is, then, applied on the BEs and the total level of maturity for the overall hazard group is calculated by aggregating the BEs' maturities. The needed data and information that supports the model development were found in the technical reports provided by EDF, which are not mentioned here for confidentiality reasons.

3.4.1. Description of the hazard groups PRAs

In this section, we consider a case study extracted from PRA models of two hazard groups, i.e.,

external flooding and internal events provided by EDF. Both PRA models were developed using the Risk Spectrum Professional software.

In all generality, *"external hazards"* refer to undesired events originating from sources outside the NPP, such as external flooding, external fires, seismic hazards etc. [82]. In this case study, we consider a particular external hazard, i.e., external flooding, that is caused by the overflow of water due to naturally induced external causes, e.g., tides, tsunamis, dam failures, snow melts, storm surges, etc. [83]. The *"external flooding"* PRA model considered in this application is a combination of event trees and fault trees that are constructed to evaluate the risk of external flooding in different water level conditions (scenarios). The total risk index of external flooding is, then, calculated by summing the risk indexes at each water level. The PRA model of external flooding is complex and has a large scale, including three operation states, thousands of BEs and several thousands of MCSs.

"*Internal events*" refer to undesired events that originate within the NPP itself and can cause initiating events that might lead to loss of important systems and, eventually, a core meltdown [19]. Major internal events include components, systems or structural failures, safety systems operation, and maintenance errors, etc. [84]. Internal events might also lead to other initiating events like turbine trip and Loss of Coolant Accidents (LOCAs). In nuclear PRA, internal events are considered a well-established and understood hazard group [36], and highly mature PRA models are available for their characterization. The internal events PRA model considered in this case study is based on a combination of event trees and fault trees that are constructed for evaluating the risk over different internal events (e.g., loss of offsite power, loss of auxiliary systems). The risk index of the entire internal events hazard group is, then, calculated by summing the risk indexes (i.e., minimal cut sets at a given operation state and scenario) of the individual internal events. Similarly to the PRA model of external flooding, the PRA model of internal events is complex and has a large scale, also containing three operation states, few thousands of BEs and several thousands of MCSs.

3.4.2. Reduced-order model construction

The first step in the developed for evaluating the level of maturity is to construct the reduced-order model. Here, we only show in details how to construct the reduced-order risk assessment model for the external flooding PRA model. For the internal events PRA model, the reduced-order model can be constructed in a similar way.

In this case study, we set the fractions of the risk to be $\alpha = \beta = \gamma = 0.8$. From Eq. (3.6), we found

that only one out of six operation states (NS/SG-normal shutdown with cooling using steam generator-NS/SG) is needed for the reduced-order model, which contributes to 86% of the total risk index. Therefore, we have $n_O = 1$. Similarly, based on Eq. (3.7), only one out of ten scenarios (water levels) is needed for the reduced-order model, whose risk contribution is 98.7%. Hence, we have $n_S = 1$. Based on Eq. (3.9), given the operation states and scenarios of interest, 5 out of 3102 MCSs already contribute to 80.1% of the risk at the given operation state and scenario. Thus, we have $n_{MCS} = 5$. Then, a reduced-order model can be constructed using the atomic elements in Table 3.4. The definitions of BEs in the MCSs of Table 3.4 can be found in Table 3.5. An illustration example on the pathway of the first minimal cut sets is given in Figure 3.6. Assuming the rare-event approximation, the risk index of interest, i.e., the probability of core meltdown, can be calculated using the MCSs and the BEs in Table 3.4, following Eqs. (3.6), (3.7), (3.9) and (3.10). The constructed reduced-order risk model can reconstruct $86\% \times 98.7\% \times 80.1\% = 67.99\%$ of the total risk R.

Table 3.4 Reduced-order model constituents

Operating state	Scenarios	MCS
NS/SG	Water level A	MCS1={BE1, BE2, BE3}
		MCS2={BE2, BE3, BE4}
		MCS3={BE3, BE5, BE6, BE7, BE8}
		MCS4={BE2, BE3, BE7, BE9}
		MCS5={ BE2, BE3, BE6, BE10}

Table 3.5 Basic events included in the reduced-order model

Symbol	Basic event
BE1	External flooding with water level A inducing a loss of offsite power
BE2	Loss of auxiliary feedwater system due to the failure to close the isolating valve
BE3	Loss of component cooling system because of clogging
BE4	Failure of all pumps of the Auxiliary feedwater (AFW) system
BE5	Failure of the turbine of the AFW system
BE6	Failure of the Diesel Generator A
BE7	Failure of the Diesel Generator B
BE8	Failure of the common diesel generator
BE9	Failure of pumps 1 and 2 of AFW system
BE10	Failure of pumps 2 and 3 of AFW system

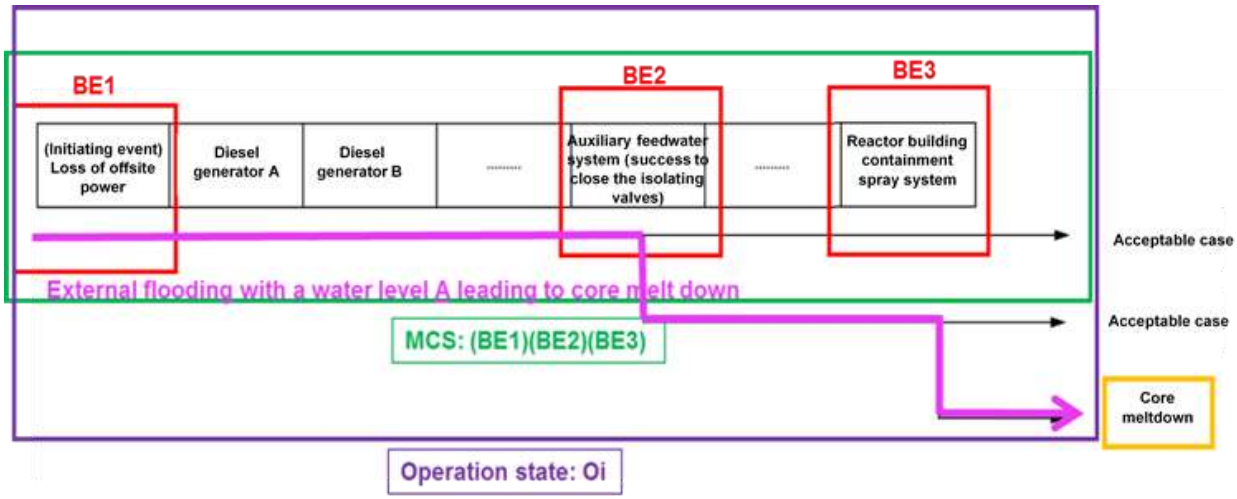


Figure 3.6 Illustration of a MCS in an individual reduced-order model

3.4.3. Evaluation of the level of maturity for external flooding hazard group

The levels of maturity for the basic events in Table 3.5 need to be evaluated using the developed method in Sect. 3.3. In the following, we illustrate in detail how to apply the developed framework on a basic event namely "External flooding with water level A inducing a loss of offsite power" (BE_1). For the other basic events, we directly give the results in Table 3.7.

As shown in Eq. (3.14), the level of maturity of a basic event is evaluated as a weighted average over the maturity attributes and sub-attributes illustrated in Figure 3.1. Hence, the weights of the maturity attributes and sub-attributes need to be determined. AHP method is adopted in this work for this purpose [48]. As illustrated in Chapter 2, two pairwise matrixes need to be constructed and filled by experts. The first is a 4×4 comparison matrix, constructed for evaluating the weights W_i (relative importance) of the attributes under level of maturity in defining their "parent" attribute i.e., level of maturity. The second is 5×5 comparison matrix constructed for comparing the weights $W_{i,j}$ (relative importance) of the strength of knowledge "daughter" attributes (i.e., sub-attributes under the strength of knowledge). For more illustration on AHP method and pairwise comparison matrixes, see Chapter 2. The results are presented in Table 3.6. Notice that, the weights are evaluated only once and used for the evaluation of all the basic events.

The next step for evaluating the level of maturity is to assess the attributes and sub-attributes presented in Figure 3.1 for BE_1 in the light of the guidelines presented in Sect. 3.2.2. In this basic event,

the probability was calculated by extrapolating the probability distributions based on observed data to the extreme water flowrate (i.e., flowrates that have never occurred). In more details, the following steps were performed:

- Water heights that lead to failure of specific equipment were defined.
- The water flowrate was predicted for the given heights at the NPP platform ensuring to cover each flowrate that can lead to the given water height at the platform.
- The flowrate was multiplied by safety factors.
- The "return period" were obtained by the same law that was used to estimate the millennial flooding flowrate of the river of interest.
- The return periods for flowrates of interest were then, calculated by extrapolating the flooding data curves toward extreme values (at low probabilities) of flow at the platform of the power plant.
- The frequencies (frequency =1/ return period) were then, rounded and mean values were obtained by the law for the flowrates of the Millennial Flood.
- The frequency of each interval is chosen to be the maximum frequency at the whole height interval.
- No uncertainty analysis was taken into account for estimation the frequencies of the critical heights.
- Due the basin special characteristics, the analysts are forced to consider the "renewal theory" (combining two statistical models of occurrence of events and their magnitude together).

Comments:

- Experts have confidence in the calculation used to convert the heights into flowrates because they are based on solid deterministic models.
- Experts have doubts on extrapolating the frequency to the extreme flowrates.
- This result is also to be considered with caution since they are based on the current limited models and knowledge.
- Multiplying the flowrates by safety and augmentation factors is considered conservative.
- The characteristics of the river basin are special in view of the evolution of the distributions of extreme floods, which opens more room for uncertainty.
- Using renewal theory-based approach is considered conservative.

- High uncertainty is presented in the analysis.

From the previous arguments, one can notice that there is uncertainty about the acknowledged limitations and implicit assumptions (unmodeled uncertainty). This meets Level 4 of uncertainty, which leads the analysts to assign a score of (2) From Table 3.1.

For the conservatism attribute, it is not possible in this case to consider the conventional acceptance criteria (e.g., acceptable core meltdown of 10^{-4}) since we are considering only one hazard group. Accordingly, experts were asked to assign an artificial value for the acceptable external flooding probability, in order to compare it to the estimated external flooding risk value of our model of interest. Now, since the analysis of the external flooding probability is based on hydrodynamic model, it is considered to be realistic but with low level of confidence. From Figure 3.3, since we are comparing the risk metric to an acceptance criteria, it was found that the conservative estimates are misinforming. A score of 2 was assigned for the conservatism.

The sensitivity of this basic event is calculated by Eq. (3.1). The basic events probability is altered by 50%. Which leads to the total change in the model output by 50% (since this basic event appears in each minimal cutset and has a Fussell-Vesely importance measure of 1). From Table 3.3, this corresponds to a level 4 of sensitivity, which in turn, corresponds to a score of 2 in the light of maturity.

The same way of reasoning was adopted for evaluating the scores of knowledge attributes. The results are shown in Table 3.6. The maturity attributes scores are then, aggregated by Eq. (3.14). The level of maturity for BE_1 is found to be 2.15.

Table 3.6 Assessment of "leaf" attributes (BE_1)

Attribute	U	C	S	K							
Sub-attribute	-	-	-	A	Co	QD	QA	Ph	Ex	AM	P	PM
W_i	0.30	0.15	0.15							0.40		
$W_{i,j}$	-	-	-	0.25	0.06	0.17	0.17	0.10	0.05	0.10	0.05	0.05
Score	2	2	1	1	5	3	2	3	5	3	5	5

The same steps are repeated for all the basic events and presented in Table 3.7. The final step before evaluating the overall level of maturity for external flooding hazard group $m_{ext-flood}$, is to determine the weights of each basic event, in a given elementary model and the corresponding elementary model by Eq. (3.12) and Eq. (3.13).

From Eq. (3.12), the weight of the elementary model is: $W_1 = \frac{R_l}{\sum_{l=1}^{1} R_l} = 1$.

From Eq. (3.13), the weight of the basic event in the given elementary model is: $W_{1,1} = \frac{I_{l,1}}{\sum_{q=1}^{n_{l,q}} I_{l,q}} =$ 0.320

The same procedure are repeated for each basic event and the results are presented in Table 3.7. Finally, the overall level of maturity for the hazard group is evaluated by Eq. 3.15. The level of maturity is found to be $m_{ext-flood} = 2.45$.

Table 3.7 Knowledge assessment and aggregation over the basic events

BE	BE1	BE2	BE3	BE4	BE5	BE6	BE7	BE8	BE9	BE10
$m_{l,q}$	2.150	1.488	2.690	3.948	4.002	4.002	4.038	3.962	3.908	3.908
$I_{l,q}$	1.000	0.9020	0.553	0.182	0.141	0.127	0.121	0.045	0.028	0.028
$W_{l,q}$	0.320	0.289	0.177	0.058	0.045	0.041	0.039	0.014	0.009	0.009

The same steps are repeated for the internal events hazard groups and the maturity was found to be $m_{internal} = 3.87$.

Finally, for risk maturity aggregation, we adopt the technique presented in Sect. 3.3.5 where the risk is represented as a risk tuple (R, M). Please note that the risk presented here after are artificial and the real number that provided by EDF are not presented for some confidentiality reasons.

External flooding risk tuple: $\left(R_{ext-flood}, m_{ext-flood}\right) = (1.5^{-5}, 2.45)$

External flooding risk tuple: $(R_{internal}, m_{internal}) = (1.2^{-7}, 3.87)$

First, by Eq. (3.16) the total risk is calculated arithmetically $R = 1.512^{-5}$. Then the level of maturity is calculated by Eq. (3.17). Two variables need to be considered, the level of maturity m_i of a given hazard group, and its corresponding weight (relative importance). The hazard group weight is calculated by Eq.(3.18) and found to be $W_{ext-flood} = 0.992$ and $W_{internal} = 0.008$. Finally, the overall maturity is found to be 2.462 and the risk tuple is $(1.512^{-5}, 2.45)$.

3.4.4. Results and discussion

As expected, the level of maturity for internal events ($m_{internal} = 3.87$) is higher than that for external flooding ($m_{ext-flood} = 2.45$). This means that the analysis and the results of the internal events are more realistic than these for external flooding. This can be explained by the fact that unlike external flooding, risk analysis for internal events hazard group in NPP has been performed for all power plants all over the world, which in turn, created the opportunity to develop solidly the appropriate models, level of details and base knowledge required for realistic evaluations [19]. This leads to a relatively well established highly mature

PRAs [36]. On the other hand, as seen in the example above: most of the risk is contributed by BE_1 (external flooding with water level A inducing a loss of offsite power), BE_2 (loss of auxiliary feedwater system due to the failure to close the isolating valve), and BE_3 (loss of component cooling system because of clogging). The three basic events probabilities are obtained based on relatively, low level of knowledge, high conservatism and high uncertainty. For example, the probability of occurrence of BE_1 is calculated by extrapolating the distributions based on observed data to the extreme water flowrate (i.e., flowrates that have never occurred). Besides, the probabilities of floods were taken as mean values without considering the uncertainty analysis. In addition, the characteristics of the river basin are special in view of the evolution of the distributions of extreme floods, which opens more room for uncertainty.

The overall risk is represented by $(R, M) = (1.512^{-5}, 2.45)$. Most of the risk and level of maturity in this tuple is on account of external flooding hazard group, which in turn, explains the low level of maturity on the overall risk.

3.5. Conclusion

In this chapter, we have proposed a method for evaluating qualitatively the different degrees of realism and maturity in risk contributor's analysis. In this framework, we tried to focus on the attributes that are believed and emphasized in the literature to affect the level of realism and maturity of analysis, and most importantly, the process of DM. The framework is based on four main attributes: uncertainty, conservatism, strength of knowledge and sensitivity. The strength of knowledge attribute, was further broken into five sub-attributes (data availability, data consistency, source of data, quality and reliability of data, experience and value-ladenness of the analysts. Analytical Hierarchy Process (AHP) is adopted to apply the framework, where pairwise comparison matrixes were built to estimate the relative weights of the attributes. An assessment protocols were developed to facilitate the process of attributes evaluation for a given problem. A technique called the reduced-order model was also developed to allow the application of the developed framework on the level of constituting elements (basic events), which in turn, leads to a more relevant and accurate assessment. Finally, the developed framework was applied on two hazard groups in NPP; namely, external flooding and internal events. The application of the framework to a case study stresses the importance of accounting for the level of maturity of a given hazard group for better informing DM. For example, the level of maturity can be very important in informing the decision maker in contexts where an option needs to be chosen, or to assess if the analysis are sufficiently mature or need to be enhanced for making a decision.

A potential limitation of the developed approach is that it was developed to be applied only on the level of “atomic elements” and not the level of the model structure. Therefore, the framework need to be enhanced in the future to consider two levels of analysis: the level of atomic elements and the level of the model structure. In addition, we do not pretend that the framework itself is complete in terms of the attributes and factors that affect the level of maturity. However, it still stands a good starting point for overcoming the heterogeneity in the maturity level of the hazards group that in turn lead to mathematical inconsistent and physically non-meaningful results. Finally, please note that it is out of the context of this work to show in details the process of DM given this maturity index.

Chapter 4 Assumptions in risk assessment models and the criticality of their deviations within the context of decision making

The trustworthiness of risk assessment models depends crucially on the validity and solidity of the assumptions made. In PRA, assumptions are typically made, based on best, conservative, or (sometimes) optimistic judgments. Best judgment and optimistic assumptions may result in failing to meet the quantitative safety objectives, whereas conservative assumptions may increase the safety margins but result in costly design or operation. In the present chapter, we develop an extended framework for evaluating the criticality (risk) of the deviations from the assumptions made in the risk assessment, which might lead to a reduction of the safety margins. In particular, a review of the approaches proposed in the literature to assess the assumptions and assumptions deviation risk is presented in Sect. 4.1. In Sect. 4.2, we present the extended method. Then, in Sect. 4.3, the implementation procedures are illustrated, and an application of the framework to a real case study of NPP is presented in Sect. 4.4. Finally, in Sect. 4.5, we offer a discussion and some conclusions.

4.1. State of the art

In risk analyses, assumptions are inevitably made by experts because of incomplete knowledge, data, information and understanding of the phenomena involved [11], for simplifying the analysis when necessary [10]. The recognition of the importance of assumption on the results of risk assessment led experts in the field to formulate some methods to evaluate the quality of assumptions and to treat the uncertain ones.

As seen from Chapter 1, the NUSAP is applied in [21], [22], [10], [23] to assess the quality of assumptions through a pedigree diagram. Also, some methods are proposed for treating uncertain assumptions [11]: (i) law of total expectation; (ii) interval probability; (iii) crude strength of knowledge and sensitivity categorization; (iv) assumption deviation risk. First, in the “law of total expectation”, a probability distribution expressing the belief on different assumptions is introduced [11]. This kind of

techniques is appropriate when the beliefs on the assumption are based on strong knowledge and historical data [11]. Second, in the "the interval probability", the assessors are asked to assign the minimum and maximum values of assumptions and their corresponding believed probability [11]. This technique is more appropriate for cases of weak knowledge [11]. Third, in the crude SoK and sensitivity categorization, the criticality of assumption is assessed by assessing the strength of knowledge on which the assumptions are made, as well as the dependency of risk assessment on this assumption [11]. Finally, for the assumption deviation method, the risk of deviation is evaluated considering three elements: the degree of expected deviation of assumption from reality, the likelihood of the deviation, and the knowledge on which the assumptions are based [17], [26]. This method was later extended in [11], where some setting were defined given the belief in the deviation from the assumption, the sensitivity of the risk index and its dependency on the assumption, and the SoK on which the assumptions are made [11]. Guidance for treatment of uncertainty related to the deviation of assumptions were given for each setting. However, the aforementioned methods either do not comply with evaluating the level of trustworthiness of risk assessment models within the context of hierarchical framework, lack of a rigorous evaluation protocols, or do not comprehensively consider different types of assumptions, e.g., conservative assumptions and DM, e.g., comparing alternatives.

4.2. The proposed framework

In this section, the original work of Khorsandi and Aven (2017) [26] is extended. Compared to previous works on the subject, we consider also conservative assumptions, other contexts of DM, and introduce decision flow diagrams to support the classification of the criticality of the assumptions made.

In this work, we assume that each assumption As_i affects the numerical values of some parameters in the Probabilistic Risk Assessment (PRA) model. The factor that links the assumptions to the numerical parameters is called "juncture" in this paper. The criticality (C) of an assumption is assessed based on the six criteria: (i) the type of assumption; (ii) the context of decision making; (iii) the belief (likelihood) in deviation from reality; (iv) the amount of deviation from reality; (v) the likelihood of the deviation; (vi) the margin of deviation; (vii) the strength of the knowledge supporting the assumption made. Three levels of criticality are defined with their corresponding settings:

1. Very critical ($C = 1$): The assumption is based on weak knowledge and the confidence on the assigned value of the model parameters is low. Besides, the assumption deviation has severe influence on the decision making and might lead to exceedance of the safety limit.

Further analysis and justification of the assumption is required.

2. Not very critical ($C = 2$): The assumption is made based on a moderate level of knowledge. The assumption deviation is likely to happen, but the risk metric remains within the safety limits even after considering such assumption deviation. The assumption can be trusted to support DM if the risks of the deviation from other assumptions are all not critical ($C = 3$). Further analysis and justification of the assumption is needed only when multiple other assumptions are also in this state.

3. Not critical ($C = 3$): The assumption made is based on strong knowledge. An assumption deviation is unlikely to happen or, if it happens, it does not affect the DM. The assumption can be trusted and decisions can be made based on the current assumption.

To evaluate the criticality of the assumptions deviations, six criteria are considered, as shown in Figure 4.2.

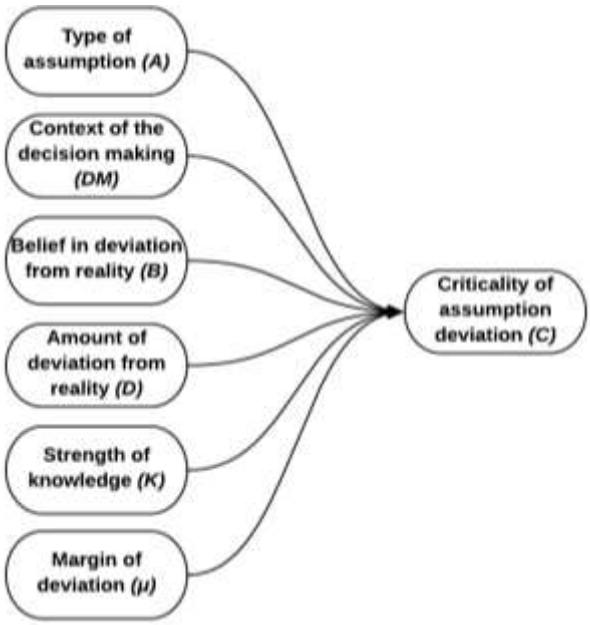


Figure 4.1 Criteria for evaluating the criticality of assumption deviation

1. Type of assumption (A): Assumptions made in PRA can be classified into different types. For example, [19] distinguishes three types of assumptions: conservative assumptions, best judgment assumptions and approximations. Conservative assumptions are made out of cautiousness and tend to overestimate the risk rather than underestimate it; best judgment assumptions are believed to represent expected scenarios, given the available knowledge; approximations are assumptions that are made for reducing the complexity of the models [20]. Deviations in different types of assumptions

might lead to different influences on the PRA. In our framework, three types of assumptions are considered:

i. Optimistic assumption (A_1): the assumption is judged by peers to underestimate the risk when compared to reality

ii. Best judgment (A_2): the assumption is judged by peers as representative of reality (realistic)

iii. Conservative assumption (A_3): the assumption is judged by peers to overestimate the risk when compared to reality (pessimistic).

2. Context of the decision making (DM): Risk metrics are used to support DM in different contexts [19]. In this work, we distinguish between two contexts of DM. First, the comparison with safety objectives, whereby the risk metrics are compared to quantitative safety goals and criteria [19]. In this case, the decision maker would accept performing the task (project, task, work, etc.) if the risk metric is lower than the safety objective; otherwise, some safety reduction measures (e.g., safety barriers, safety systems, etc.) need to be implemented in order to reduce the risk. Second, the comparison of alternatives, whereby risk metrics of different alternatives are compared. In this case, the decision maker would choose the alternative that leads to a lower risk, or choose the risk reduction measure that leads to a higher reduction of the risk metric given the cost of the application. The criticality of assumptions deviations varies from one context to another, where, in comparing risk metric to a safety goal, only the deviation toward critical scenarios need to be considered. On the other hand, for comparing alternatives in terms of their risks, all the deviation scenarios need to be considered, since a conservative assumption might lead to a higher risk metric and hence, lead the decision maker to make a wrong decision by choosing another alternative that has a higher risk in reality but appears lower due to the different levels of conservatism in the analysis.

3. Belief in deviation (B) measures the realism of an assumption and is expressed by the likelihood of assumption deviation. The likelihood is assigned by the experts following the criteria defined in [26], i.e., what could cause the assumption to deviate in reality; what are the key drivers of those causes; etc.

4. Amount of deviation from reality (D) refers to the amount of deviation between the assumed parameter value and the true value. It is assigned by experts and expressed in percentage.

5. Strength of knowledge (K) refers to the strength of the background knowledge that supports the evaluation of the belief in deviation and the amount of deviation.

6. Margin of deviation (μ) refers to the degree to which an assumption may deviate before the deviation changes the decisions made based on the results of risk assessment, e.g., the violation of the acceptance criteria or the change of the prioritization of different options. This margin is calculated analytically (see Sect. 4.3.8) and expressed in percentage.

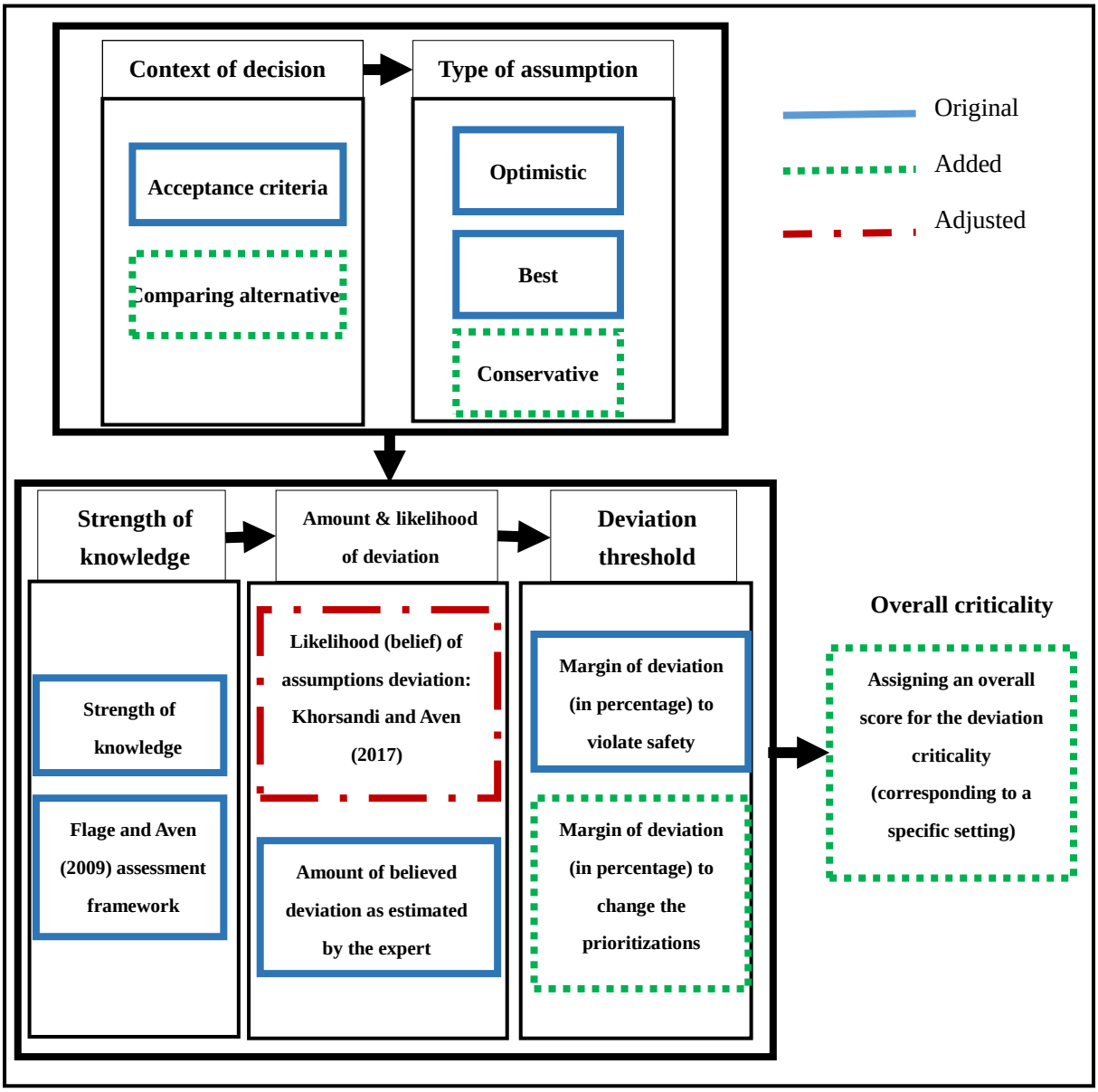


Figure 4.2 A comparison between the original (Khorsandi & Aven, 2017 [26]) and the extended frameworks for assumption deviation risk assessment

The logical combination of the six criteria yields different levels of criticality. Decision flow diagrams are introduced in this work to capture the logical relationship between the six criteria and the criticality of assumptions deviations. Only one example on decision flow diagram is presented in Sect. 4.3.9, the rest of the decision flow diagrams are presented in the appended paper III (Sect. 2.2.9).

A comparison between the original assessment framework in Khorsandi and Aven (2017) and the extended framework is made in Figure 4.2. It can be seen that the original work in [26] is adjusted and extended to include an additional context of DM (comparing alternatives) and also a new type of assumption (conservative assumptions). Accordingly, new criteria are added or adjusted to integrate the new decision context and type of assumption in the assessment of the assumption deviation risk. As to the presentation of the assumption deviation risk, the radar plot in [26], which presents the contributing factors to the assumption deviation risk individually, is replaced with an overall integrated metric for assumption deviation risk, i.e., the criticality (C). These extensions make it possible for the extended framework to provide a more comprehensive description of the risk from assumptions deviations.

4.3. Implementation of the framework

As shown in Figure 4.3, nine main steps are needed for applying the developed framework to assess the criticality of assumptions deviations. The nine steps are discussed in details in sub Sect. 4.3.1-4.3.9.

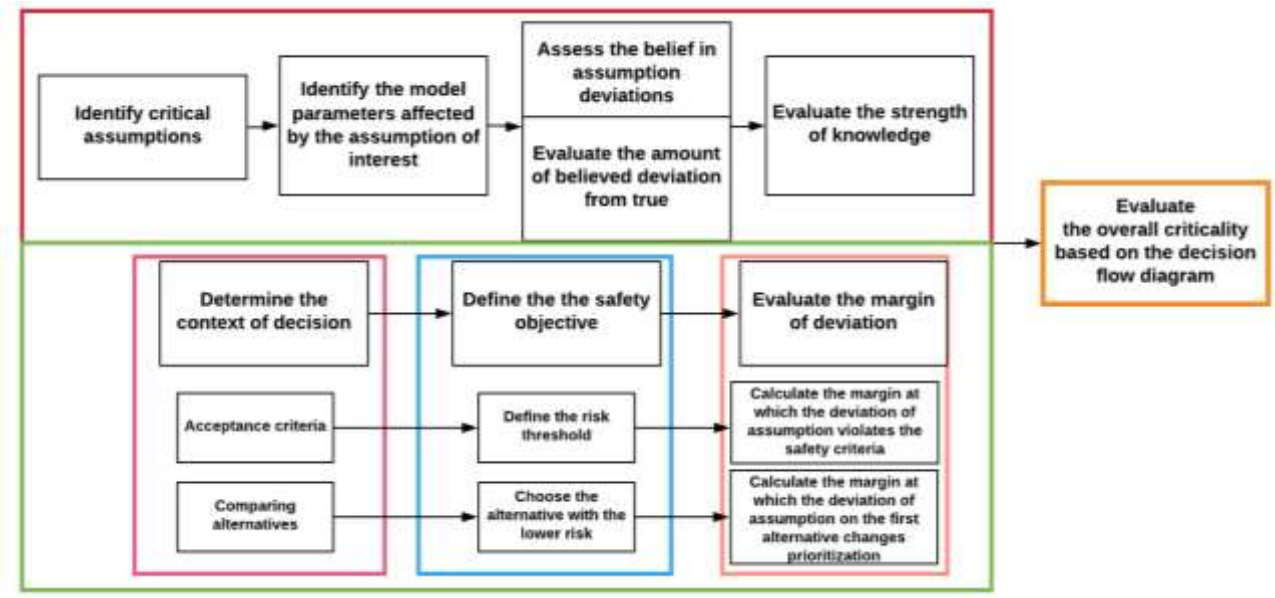


Figure 4.3 Procedure for applying the developed framework for assumption deviation criticality (risk) assessment.

4.3.1. Identify critical assumptions

In the first step, the assumptions made in the PRA are identified. The assumptions might be made due

to lack of understanding and knowledge about a phenomenon or as an attempt to reduce the modeling details and complexity [20], [19]. The type of each assumption (A) is determined by expert judgment, making reference to the definitions in Sect 4.2.1.

4.3.2. Identify the model parameters affected by the assumption of interest

As mentioned in Sect 4.2, in this work, we assume that there is a juncture that connects numerically an assumption to one or more parameters in the PRA model. Without losing generality, let us assume that the PRA model is represented by:

$$R = f(p_1, p_2, \dots p_m, \dots p_n), \tag{4.1}$$

where R is the risk metric and $p_1, p_2, \dots p_m, \dots p_n$ are the model parameters (e.g., failure probabilities), f is the function that depends on the structure of the model. where As represents a set of assumptions. In the framework, we only consider the assumptions that can be altered numerically or that can change the numerical values of the model parameters. We do not consider the assumptions that are related to the model structure or that cannot be measured numerically. The second step, then, involves identifying the model parameters affected by each assumption, as shown in Figure 4.4.

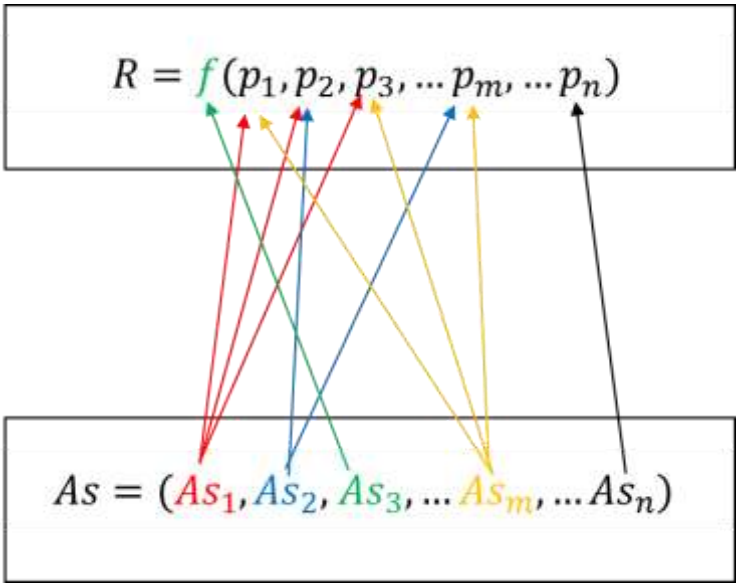


Figure 4.4 Representation of connections between assumptions and model parameters

4.3.3. Assess the belief in assumption deviation

The belief in deviation is evaluated as the subjective probability assigned by experts that the assumption deviates from the actual conditions. The assigned value is conditional on the available background knowledge, including experts' individual expertise. It should be noted that the aim of evaluating the belief in deviation is not to assign a precise value for the probability of deviation. Rather, it

aims at expressing the experts' beliefs , based on the available knowledge, on how likely the assumption might be deviating from reality [26]. Such a step can be regarded as a tool for making good use of experts' individual expertise by reflecting their implicit knowledge that cannot be directly stated or documented.

To determine the value of B, the likelihood (l) needs to be evaluated by experts first, following the considerations recommended by Khorsandi and Aven (2017) [26]: What could cause the assumption to deviate? What are the key drivers of those causes? Has a similar deviation occurred in the past? What evidence is available for supporting the potential for a deviation?

Then, the value of B is determined based on the likelihood (l):

a. $B = 1, if\ 0 \leq l \leq 20\%$

b. $B = 2, if\ 20\% < l \leq 30\%$

c. $B = 3, if\ 30\% < l \leq 100\%$

4.3.4. Evaluate the amount of believed deviation from the true value

The amount of believed deviation is evaluated as the relative distance between the assumed parameter value and the true value believed by experts, as expressed by Eq. (4.2). Similar to the belief in deviation, the believed deviation D is evaluated by experts and represents the experts' belief on how severe the deviation could be. The value assigned to D takes a positive sign $(+)$ if the assumption is believed to deviate towards dangerous scenarios and a negative sign $(-)$ if it is deviating towards safe scenarios:

$$D = \frac{p_t - p}{p} \tag{4.2}$$

where D is the amount of believed deviation, p_t is the parameter value believed true by the experts, and p is the parameter value as assumed in the analysis.

4.3.5. Evaluate the strength of knowledge

The assigned belief (likelihood) and amount of deviation are conditional on the background knowledge available, and on the individual expertise and points of view of the experts who made the assessment. Therefore, the strength of knowledge on which the assessment is based is highly relevant and is explicitly considered in both the original and extended framework. In this work, we use the method proposed in [6] for evaluating the strength of knowledge. This approach is mainly based on the evaluation of four criteria: (i) reasonability and realism of assumptions; (ii) phenomenological understanding; (iii) availability of reliable data and information; (iv) agreements among peers. In addition, we take into account

a fifth criteria, suggested by Khorsandi and Aven (2017): (v) the level of expertise and competence of the experts. A score of 1-3 is given for each criterion, corresponding to three levels, i.e., weak, moderate and strong, respectively.

A weighted average of the five criteria scores $k_i, i = 1,2, \dots 5,$ is used to calculate the overall knowledge score SK:

$$SK = \sum_{i=1}^{5} w_i \cdot k_i, \tag{4.3}$$

where w_i is the weight of criterion k_i. Obviously, the five criteria are not equally important in defining the strength of knowledge. To handle this, the AHP [48] is used to determine the weights of the strength of knowledge criteria. More illustration on AHP method and how to apply it is presented in chapter 2. The strength of knowledge denoted by K, is, then, calculated based on the value of SK:

- $K = 1,$ if $1 \leq SK \leq 1.6$
- $K = 2,$ if $1.6 < SK \leq 2.3$
- $K = 3,$ if $SK > 2.3$

4.3.6. Determine the context of decision

In the original work in [26], only one context of DM was considered, i.e., comparing a risk metric to a specific safety objective. In this sense, only assumptions deviations toward dangerous scenarios (optimistic assumptions) need to be considered. However, in the practice of risk management, we often need to compare alternatives in terms of their risks (e.g., two options leading to risks or choosing among two options implemented to reduce the risk). In this case, all the deviation scenarios need to be considered, since a conservative assumption might lead to an "unrealistically" higher risk metric, which, in turn, leads the decision maker to prefer the alternative with the "unrealistically" lower risk metric; in other words, it gives a "false alarm" of high risk. For more illustration, take the example in Figure 4.5. In this example, the decision maker is comparing two alternatives, Al_1 and Al_2, and he/she prefers to choose the alternative with the lower risk. At a first glance and using conservative assumptions, the decision maker would choose Al_1 as it has a lower risk metric value (the blue solid line). However, a second look shows that the value of R_2 (in the meshed filling) is lower than that of R_1, when the true condition is used in the calculation rather than a conservative assumption. Hence, it is important to identify the context of DM when implementing the extended framework. In this work, two DM contexts are distinguished, namely, comparing a risk metric to a safety objective (DM_1) and comparing two alternatives (DM_2).

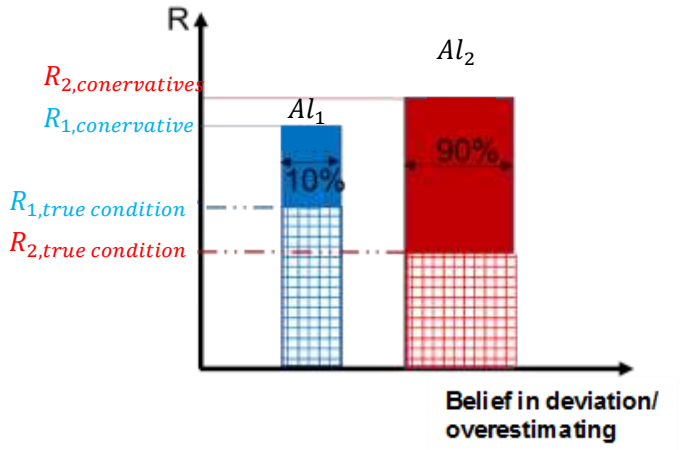


Figure 4.5 Comparing the risk related to two alternatives taking into account the risk metric value based on the assumption made and the true condition

4.3.7. Define the safety objective

The safety objective needs to be identified considering the given decision context, as shown in Figure 4.3. The safety objective represents a numerical value whose exceedance by the risk metric would lead to changes in the results of the risk-informed decision making. The safety objective is dependent on the context of the DM. For the decision context DM_1, the safety objective is identified as the threshold that the risk metric should not exceed. On the other hand, if the decision context is DM_2, the assessor needs to choose the alternative with the lowest risk metric value. Therefore, the (higher) risk metric value of another alternative is defined as the safety objective under this DM context.

4.3.8. Identify the margin of deviation

Next, the margin of deviation (μ) needs to be calculated. This margin represents the maximum tolerable assumption deviation before the risk-informed decision is changed. As shown in Figure 4.4, different assumptions might affect one or more model parameters, or, the other way around, a model parameter might be affected by one or more assumptions. In this work, we calculate the margin of deviation one assumption at a time, to reduce the complexity of the analysis. Assume that the assumption of interest a_i affects model parameters $p_1,\ p_2, \dots p_m$. Then, we assume that the assumption deviation affects "similarly" the related parameters $(p_1,\ p_2, \dots p_m)$ to make the equation solvable. The assumption deviation can be modeled by:

$$\begin{cases} \acute{p_1} = (1+\mu)p_1 \\ \acute{p_2} = (1+\mu)p_2 \\ \vdots \\ \acute{p_m} = (1+\mu)p_m \end{cases} \tag{4.4}$$

where $\acute{p_i},\ i = 1, 2, \dots m,$ are the deviated model parameters and μ represents the amount of deviation in

the model parameters (and assumed to be the same for all parameters affected by an assumption) due to the deviation in the assumption. It should be noted that in theory, the basic event probabilities can also change by different amounts, resulting in different values of μ for different basic events. Then, the deviated risk metric $\acute{R}$ is calculated by:

$$\acute{R} = f(p_1', p_2', \dots p_m', p_{m+1} \dots p_n) \tag{4.5}$$

The value of μ can be calculated by solving the following equation:

$$\underset{\mu}{arg}\, f\left((1+\mu)\cdot p_1, (1+\mu)\cdot p_2, \dots (1+\mu)\cdot p_m,\ p_{m+1}, \dots p_n\right) = R_{th} \tag{4.6}$$

In Eq. (4.6), R_{th} is the safety objective defined in Sect. 4.3.7, i.e.:

$$R_{th} = \begin{cases} R_{lim}, if\ the\ decision\ context\ is\ DM_1 \\ R_2, if\ the\ decision\ context\ is\ DM_2 \end{cases} \tag{4.7}$$

where R_{lim} and R_2 represent the safety limit objective and the risk metric value of the alternative being compared, respectively.

4.3.9. Evaluate the overall criticality based on the decision flow diagrams

The criticality of an assumption deviation measures its influence on the risk-informed decision making and, hence, on the safety of the system. As defined in Sect. 4.2, the criticality of the assumption deviation depends on both the severity of the influence and the likelihood of the deviation. Four scenarios are distinguished to quantify the severity of the influence of the assumption deviation:

a. failures in meeting the established objectives, i.e., the magnitude of deviation is larger the deviation margin, leading to the exceedance of the safety limit;
b. success in meeting the established objectives i.e., the magnitude of deviation is lower than the deviation margin, or the deviation is occurring towards lower amounts of risk due to conservatism in the assumption;
c. Altering the different prioritization when comparing two or more alternatives, i.e., the risk metric based on unrealistic assumptions is higher or lower than what it would be based on the true conditions, leading to the mischoice among the different alternatives.
d. Unchanging the prioritization when comparing two or more alternatives, i.e., the risk metric based on unrealistic assumptions is higher or lower than what it would be based on the true conditions, leading to misranking the different alternatives.

Considering the scenarios defined above and the likelihood of deviation, decision flow diagrams are built for evaluating the criticality of assumption deviation risk. We present only one example on the

decision flaw diagram In Figure 4.6, the rest are presented in the appended paper III (Sect. 2.2.9). It should be noted that in these figures, the difference between the margin of deviation μ and the amount of deviation D, denoted by $\Delta\mu$, is calculated and used to measure the safety margin for a given assumption deviation:

$$\Delta\mu = \mu - D \tag{4.8}$$

Following the steps in Sects. 4.3.1-4.3.8, the criticality C can be evaluated using the decision flow diagrams presented in Figure 4.6 and appended paper III (Figure 6-8).

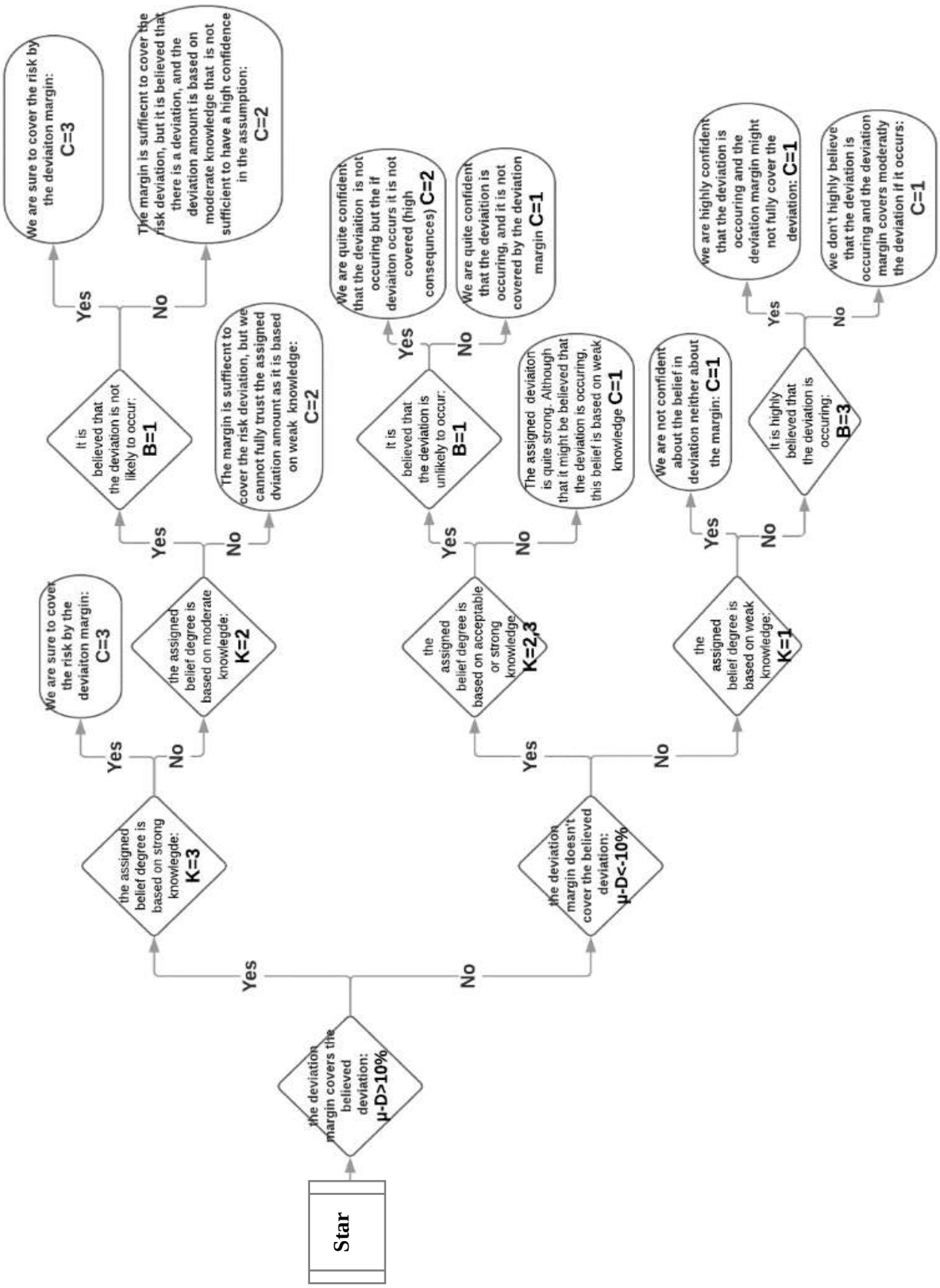


Figure 4.7 Criticality assessment decision flow diagram for decision context DM_1 and assumptions of types A_1 and A_2.

4.4. Application

In this section, we apply the developed framework on real PRA models for the external flooding hazard groups of the same case study presented in Chapter 3. The PRA model for external flooding is chosen because it is less mature compared to the PRA model of other hazard groups and involves many assumptions.

4.4.1. Evaluation of assumption deviation risk

4.4.1.1. Identifying critical assumptions

The critical assumptions in the PRA model of external flooding (the basic events of the external flooding listed in Table 3.4), are identified following the procedures in Sect. 4.3.1 and listed in Table 4.1. The assumption deviation risks for the assumptions in Table 4.1 need to be evaluated using the developed method in Sect 4.3. In the following, we illustrate in detail how to apply the developed framework on one conservative assumption, namely "the clogging accompanying some floods is unpredictable and unfilterable". For the other assumptions, we directly give the classification results in Sect. 4.4.2.8.

Table 4.1 List of the assumptions related to the reduced-order model of the external flooding hazard group.

As_i	Description	Type	Affected basic event
As_1	It is assumed that failure to close the isolating valves for volumetric protection sealing-water proofing causes the total loss of Emergency Feed Water System (EFWS)	Conservative	BE2
As_2	If the floods occur, the clogging is certain $(P = 1)$	Best judgment	BE3
As_3	If the river flooding is accompanied with clogging, then, it is unpredictable and unfilterable	Conservative	BE3, BE4
As_4	Clogging leads to failure of Essential Services Water System (component cooling system) and therefore, the reactor containment spray system	Best judgment	BE3, BE4
As_5	It is assumed that probabilities of a given level of flood can be calculated by extrapolating the distributions based on observed data to the extreme water flowrate (i.e., flowrates that have never occurred) and that the probabilities of floods can be taken as mean values	Best judgment	BE1
As_6	It is assumed that once the water reaches the bottom of an equipment, the equipment fails	Conservative	BE2-BE 10
As_7	It is assumed that once the water level exceeds the height of the barriers, the water will enter and fill the building	Best judgment	BE2-BE 10
As_8	It is assumed that unit 1 cannot get help from unit 2 and vice versa, or from the safeguard system shared between the two units	Conservative	BE8

As_9	It is assumed that the river flood can be predicted using statistical models	Optimistic	BE1
$As_{1(}$	It assumed that once the river flood is predicted, the probability of failing to transit into the state of ""emergency shutdown" (i.e., normal shutdown and cooling with steam generator, normal shutdown and cooling with residual heat removal system etc.) is the intrinsic failure probability that is considered in normal cases	Best judgment	BE1

4.4.1.2. Identification of model parameters affected by the assumption of interest

The model parameters in the PRA model are the probabilities of the basic events in the event tree. As the clogging can lead to the loss of component cooling system (CCS) or the loss of the pumps in the auxiliary feedwater system, the assumption As_3 is related to the two basic events BE3 and BE4, as presented in Table 4.1.

4.4.1.3. Assessment of the belief in deviation

Experts from EDF are invited to assess the belief in deviation. In this assumption, the probability that the clogging is not detected and filtered is 1 ($P = 1$), while in reality, the clogging is usually detectable and can be filtered, which means that the true value of this probability is less than 1 ($P < 1$), leading to a lower risk than the value calculated using the assumed model parameters. Therefore, the experts think that this assumption is very conservative, indicating that the assumption deviation might reduce the value of the risk metric.

Some observations can also help the expert to better understand the assumption and evaluate the belief in deviation, as shown in Table 4.2.

Table 4.2 Assessment of the belief in deviation As_3

Aspects	Assessment
What could cause the assumption to deviate?	The amount of precipitation can usually be predicted. Hence, if the river flooding is caused by precipitation, then, it can be predicted. Unless it is due to barrier rupture, the river level usually increases gradually and can be seen and noticed easily. If there is heavy precipitation, the operators would pay more attention to the water filters on the river and clean the filters to make sure that the water intake is not clogged.
What are the key drivers of those causes?	The fact that the river level increases is a gradual process. The fact that the operators are able to clean the clogging if it occurs.
Has a similar deviation occurred in the past?	Yes.
What evidence is available for supporting the potential for a deviation?	The feedback reports show that a clogging has occurred before and that operators were able to see it and manage it.

Based on the analysis illustrated in Table 4.2, the belief in deviation was assigned to be 70%.

Therefore, we have $B = 3$.

4.4.1.4. Evaluate the amount of believed deviation from the true value

Experts in EDF are asked to evaluate, based on their beliefs, the amount of assumption deviation from the true values. The experts have assigned the amount of deviation in percentage to be $D = -50\%$, meaning that the experts believe that the assumption is conservative and deviating towards a higher risk.

4.4.1.5. Evaluate the strength of knowledge

The strength of knowledge has been evaluated as indicated in Sect. 4.3.5. The strength of knowledge attributes are evaluated separately, as shown in Table 4.3.

Table 4.3 Strength of knowledge criteria and weights.

Attribute	Weight	Score
Reasonability and realism of assumptions (k_1)	0.13	1
Availability of reliable data and information (k_2)	0.13	2
Phenomenological understanding (k_3)	0.42	1
Agreement among peers (k_4)	0.16	1
Level of expertise and competence of the experts (k_5)	0.16	2

The overall knowledge score K is calculated using Eq. (4.3):

$$K = \sum_{i=1}^{5} w_i \cdot k_i = 1.29$$

Then, based on the criteria defined in Sect. 4.3.5, we have $K = 1$.

4.4.1.6. Determine the context of decision making and define the safety objective

The context of the DM in this case study is to compare a risk metric to a safety limit. The risk limit for core meltdown varies between 1×10^{-5} and 1×10^{-4} [85]. As the flooding events are usually site-specific [86], the contribution of the external flooding hazard group to core meltdown also varies from one NPP to another. Moreover, we consider only a part of the external flooding PRA model in this case study (through the reduced-order model). Accordingly, for illustration purposes, we artificially set the safety limit of the considered PRA model to be $R_{lim} = 1.6 \times 10^{-8}$.

4.4.1.7. Identify the margin of deviation

As the assumption As_3 affects the basic events BE_3, BE_4, the vector of basic events' probabilities related to the assumption are $\mathrm{P}_m = (p_{BE_3}, p_{BE_4})$. Accordingly, the deviated risk function can be expressed using Eq. (4.5):

$$\begin{aligned} R^{\acute{}} = R_{th} = R_{lim} &= f\left(p_1, p_2, \acute{p}_{BE_3}, p_{\acute{BE}_4}, p_5, \dots p_{10}\right) \\ &= f(p_1, p_2, (1+\mu)\cdot p_3, (1+\mu)\cdot p_4, p_5 \dots p_{10}) \end{aligned}$$

The solver in Microsoft Excel is used to solve Eq. (4.6), with $R_{lim} = 1.603 \times 10^{-8}$. The resulted margin of deviation is $\mu_{As_3} = 26.40\%$. The margins of deviation for the remaining assumptions are calculated in a similar way, as presented in Table 4.4 next in Sect. 4.4.2.8.

4.4.1.8. Evaluate the overall criticality based on the decision flow diagram

As illustrated in Sect. 4.3, the overall criticality of assumptions deviation is assigned based on the decision flow diagrams (presented in Figures 6-8, appended paper III). For the assumption of interest (As_3), the belief (likelihood) in the deviation is assigned to be 70% (level 3). The difference between the deviation margin and the amount of believed deviation is 76.40%. The strength of knowledge is assessed to be $K = 1$. For an acceptance-criteria decision-context, this means that we believe that we are under the safety limit, and the deviation is not considered critical and can be accepted. On the other hand, our belief is based on weak knowledge, which makes it less credible. Following the decision flow diagram in Figure 4.6, the criticality of this assumption is $C = 2$. Accordingly, the assumption is not very critical and listed in the "waiting list", which means that it is accepted unless there are other criteria and information on other assumptions deviations that change the evaluation.

The same steps are repeated for each assumption. The scores and the evaluation corresponding to each criterion for each assumption are presented in Table 4.4 together with their final criticality scores.

Table 4.4 Assumption-deviation criticality and criticality criteria assessment

A_i	Type	BEs	$l_i : B_i$	D_i	μ_i	$\Delta\mu_i$	K_i	C_i
1	Conservative	BE2	95%:3	-90%	∞	∞	1	2
2	Best judgment	BE3	30%:2	90%	35.11%	-54.89%	2	1
3	Conservative	BE3, BE4	70%:3	-90%	26.40%	116.40%	1	2
4	Best judgment	BE3, BE4	5%:1	5%	26.40%	21.40%	3	3
5	Best judgment	BE1	50%:3	50%	24.22%	-25.78%	3	1
6	Conservative	BE2-BE10	90%:3	-70%	20.38%	90.38%	1	2
7	Best judgment	BE2-BE10	40%:3	30%	20.38%	-9.62%	2	1
8	Conservative	BE8	20%:1	-30%	869.95%	899.95%	1	2
9	Optimistic	BE1	40%:3	30%	24.22%	-5.78%	2	1
10	Best judgment	BE1	5%:1	5%	24.22%	19.22%	3	3

As shown in Table 4.4, the different assumptions have three levels of criticality i.e., 1, 2, 3 (very critical; not very critical; not critical). The corresponding actions that need to be taken by decision-makers and analysts are respectively:

(i) $C = 3$: The deviation is very likely to happen. Besides, the assumption deviation has severe

influence on the decision making and might lead to exceedance of the safety limit. Further analysis and justification of the assumption is required. This kind of assumptions decreases greatly the safety margin of the NPP. Therefore, it should be treated carefully.

(ii) $C = 2$: The assumption can be trusted to support decision making if the risks of the deviation from other assumptions are all not critical ($C = 3$). Further analysis and justification of the assumption is needed only when other assumptions are also in this state. This kind of assumptions does not decrease the safety margin of the NPP if the other assumptions are of the same type or less critical.

(iii) $C = 1$: An assumption deviation is unlikely to happen or, if it happens, it does not affect the decision making nor the safety of the NPP. The assumption can be trusted and decisions can be made based on the current assumption. This assumption does not impact the safety margin of the NPP.

As shown from the example above, the assumptions deviations might be inevitable. Since they might significantly affect the results of QRA, the decision makers and analysts should pay attention to their criticality. In the NPP industry in particular, some deviations might be very critical and lead to catastrophic consequences.

4.5. Conclusion

In this work, we have extended the approach of Khorsandi and Aven (2017) for evaluating assumptions deviations in QRAs. The extended framework covers a new context of DM very relevant in practice, namely, that of comparing alternatives (rather than comparing a single alternative against a safety objective) and an additional type of assumptions, namely, conservative assumptions (rather than just the best judgment type of assumptions). An integrated metric, the criticality of assumption deviation, is defined and evaluated based on the extended framework through the use of decision flow diagrams. The developed framework is applied to a case study of a PRA model of the external flooding hazard group of an NPP. The implementation of the framework has shown its feasibility and its ability to cover different types of assumptions and to provide a more complete evaluation of the assumption deviation.

The use of decision flow diagrams has both pros and cons. The pros are that these diagrams facilitate a standardized assumption deviation risk assessment, increasing both the transparency and efficiency of the assessment. These are desirable attributes in case of peer review of the assessment and considering the large number of assumptions typically involved in PRAs. A con of such diagrams are that they give a

“mechanical” assessment procedure where the assessment is based on strict rules rather than the use of overall judgments. Another possible limitation of the current research that need to be addressed in the future is that it analyzes the deviation risk for one assumption at a time and, thus, fails to take into account the deviation risk for several assumptions simultaneously.

Chapter 5 Strength of knowledge supporting risk analysis: assessment framework

In PRA, models are developed to calculate probabilistic indexes for risk characterization [6]. The outcomes are inevitably conditioned on the knowledge of the problem. Then, it is well-accepted that epistemic uncertainty must be quantified for a comprehensive characterization of risk. This relates to the Strength of Knowledge (SoK) that supports the risk modeling and assessment [15], [16]. The SoK has been identified also in Chapter 2 as a crucial part of the trustworthiness of the risk assessment outcomes.

The aim of this chapter is to develop a framework for assessing the SoK of PRA models and that can be applied on the constituting elements of a PRA model. A hierarchical framework is developed to conceptually describe the SoK and relate it to its major contributors. Sect. 5.1 briefly presents some common methods for evaluating the SoK of a risk assessment model. In Sect. 5.2, a SoK assessment hierarchical framework is developed. In Sect. 5.3, the framework is implemented in a top-down and bottom-up fashion for practical SoK assessment, based on the reduced order model presented in Chapter 3. In Sect. 5.4, a case study concerning two hazard-group PRA models of a NPP is presented. Finally, a discussion and conclusion on the method are presented.

5.1. State of the art

Few methods are found in the literature for assessing the SoK supporting risk assessment. In [6], a "crude" qualitative, direct grading of the SoK that supports risk assessment is introduced. In this method, the SoK is classified to minor, moderate, and significant with respect to four criteria: the phenomenological understanding of the problem and availability of precise and well-understood predicting models for the physical phenomena of interest, the availability of reliable data, the reasonability of assumptions made, and the agreement among experts [6], [11], [17], [41], [7]. In [17] a semi-quantitative approach known as assumption deviation risk has been introduced. The core idea of this method that poor assumptions are main sources of weak knowledge and, thus, the solidity of assumptions on which risk analysis is based should be evaluated [17], [11]. This approach is based on converting the main assumptions into uncertainty factors and identifying the criticality of assumptions by assigning crude risk scores for the main assumptions of the

risk assessment model based on: (i) the possible deviation from the assumptions and the associated consequences; (ii) the uncertainty of this deviation; (iii) the background knowledge that supports the assumptions. Similarly, [11] defines guidelines to treat the uncertainty associated with six typical settings that correspond to different levels of assumptions deviations. However, most of the aforementioned lack of an integrated framework that covers the different contributing factors to SoK. Also, they evaluate the SoK by directly scoring of some intangible contributing factors, which is hard to apply in practice.

5.2. A hierarchical framework for SoK assessment

In this section, we construct a conceptual framework to describe the SoK that supports a PRA. The framework developed, based on the review presented in the appended paper IV. The main attributes that contribute to the SoK are identified from the literature and organized hierarchically based on the framework proposed in [6], but adjusted and expanded to include more contributors and facilitate the practical implementations.

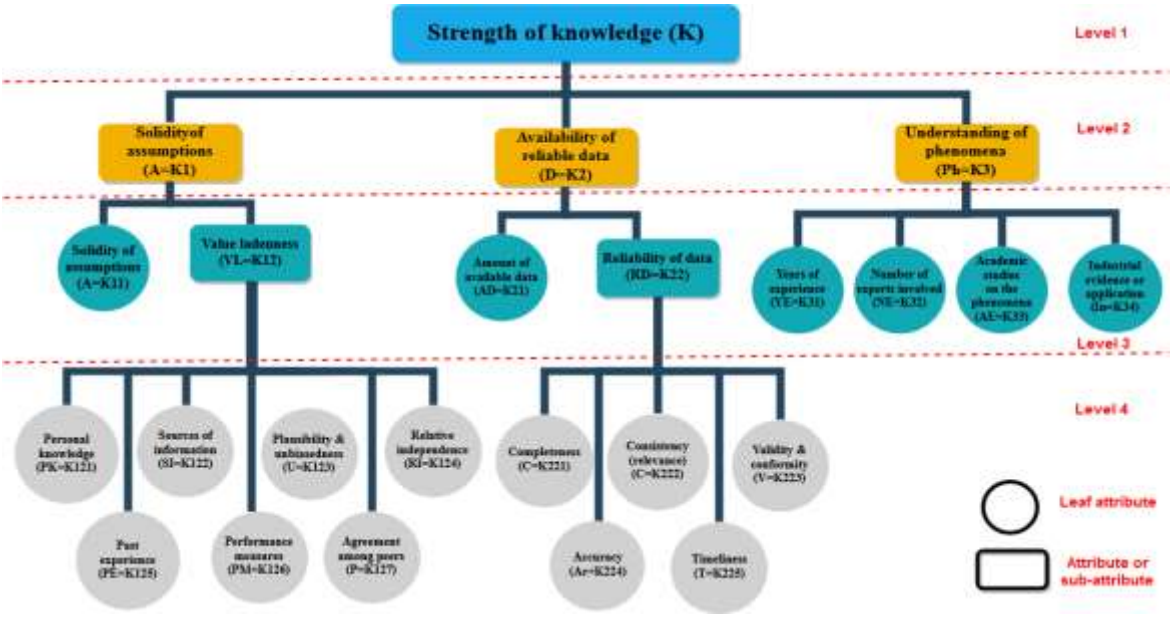


Figure 5.1 A hierarchical conceptual framework for knowledge assessment

As shown in Figure 5.1, the SoK, denoted by K (Level 1), represents the solidity of background knowledge that supports a risk model. A high value of K indicates that the model is well supported and, therefore, its results are trustable. The SoK is characterized by three level-2 attributes: solidity of assumptions (A), availability and reliability of data (D), and understanding of the phenomena (Ph). The attribute A measures the plausibility, objectivity and sensitivity of the assumptions upon which the model is based; D measures the amount and reliability of data that support the model evaluation; and Ph measures

the degree of comprehension of the phenomena involved in the risk assessment.

The three attributes of level-2 are further decomposed into sub-attributes (Levels 3 and 4) to assist their evaluation in practice. Please note that the breaking-down is designed in such a way that the sub-attributes in the same hierarchy are independent and mutually exclusive. Detailed definitions of the attributes are given in Table 5.1 and Table 5.2 Detailed guidelines for the evaluation of the attributes at the bottom levels of the framework are defined in Appendices A-C of appended paper IV.

Table 5.1 Definition of SoK attributes (Level 3)

Attribute	**Definition**
Value ladenness of the analyst ($VL = K_{12}$)	The degree to which the presumed values and beliefs that are taken as facts, and the assumptions made by experts are affected by the personal points of view, bias, subjectivity, and external or personal limitations
The sensitivity of assumption ($S = K_{13}$)	The degree to which the models' output varies with assumptions
Amount of available data ($AD = K_{21}$)	The quantity of data that supports the modeling and analysis
Reliability of data ($RD = K_{22}$)	The degree to which the available data is complete, accurate and error-free, consistent, valid and representative of reality
Years of experience ($YE = K_{31}$)	The amount of experience (measured in years) regarding a specific phenomenon
Number of experts involved ($NE = K_{32}$)	The number of experts who are explicitly or implicitly involved in understanding the phenomena and the risk analysis
Academic studies on the phenomena ($AE = K_{33}$)	The number of academic resources, i.e., articles, books, etc., available in relation to the phenomena of interest
Industrial evidence and applications on the phenomena ($IE = K_{34}$)	The number of industrial applications and reports related to the specific phenomena or events of interest

Table 5.2 Definition of SoK attributes (Level 4)

Attribute	**Definition**
Personal knowledge ($PK = K_{121}$)	The level of analysts' knowledge and relevance to the problem
Source of information ($SI = K_{122}$)	The degree of solidity, relevance, and confidence of the experts' source of information and knowledge
Unbiasedness and plausibility ($U = K_{123}$)	The experts' degree of objectivity and unbiasedness towards personal interest, or an intentional or non-intentional tendency towards a specific subject in the analysis
Relative independence ($RI = K_{124}$)	The degree of independence of the analysts from limitations or external pressures
Past experience ($PE = K_{125}$)	The experts' degree of experience in the related domain and more specifically, in the specific problem under analysis
Performance measures ($PM = K_{126}$)	The experts' degree of professionalism, skills, and competencies, past fulfillment of assigned missions and level of achievement
Agreement among peers ($P = K_{127}$)	The degree to which the assumptions made by different experts are consistent

Completeness $(C = K_{221})$	The degree to which the collected data contains the needed information for the risk modeling and assessment
Consistency $(Co = K_{222})$	The degree of homogeneity of data from different data sources
Validity $(V = K_{223})$	The degree to which the data are collected from a standard collection process and satisfy the syntax of its definition (documentation related)
Accuracy and conformity $(Ac = K_{224})$	The degree to which data correctly reflects the reality about an object or event
Timeliness $(T = K_{225})$	The degree to which data are up-to-date and represent reality for the required point in time

5.3. A top-down bottom-up method for SoK assessment

In this section, we present a top-down bottom-up method to facilitate the practical implementation of the framework proposed in Figure 5.1 for the evaluation of the SoK supporting risk assessment models. In Sect. 5.3.1, we give an overview of the SoK assessment method and how to evaluate the SoK on the level of basic elements of a PRA model. In Sect. 5.3.2, we show and how to aggregate the SoK of the basic elements to evaluate the SoK of the total risk assessment model.

5.3.1. SoK assessment for the basic events

Similar to the assessment of maturity presented in Chapter 3, the assessment of SoK starts from determining the SoK for each basic event. The total SoK for the reduced PRA model is evaluated as a weighted average of the BEs' SoK, as will be illustrated later in Sect. 5.3.2. The first step is, hence, to construct the reduced-order PRA model using the same procedural steps illustrated in Chapter 3.

After constructing the reduced order model and identifying the basic events that need to be assessed, the SoK is then, evaluated for a single basic event as a weighted average of the attributes scores presented in Figure 5.1, where the attribute scores are evaluated based on the scoring guidelines presented in Appendices A-C of the appended paper IV, which, in turn are derived based on technical reports, literature and experts' elicitation. The SoK is, then, assessed as follows:

$$K = \sum_{i=1}^{n_i} \sum_{j=1}^{n_{ij}} \sum_{k=1}^{n_{ijk}} W_i \cdot W_{ij} \cdot W_{ijk} \cdot K_{ijk}, \quad (5.1)$$

In Eq. (5.1), W_i, W_{ij} and W_{ijk} are respectively the weights of the 2nd, 3rd and 4th level attributes in the hierarchical tree of Figure 5.1, K_{ijk} is the score of the "leaf" attributes, while n_i, n_{ij} and n_{ijk} are respectively the number of attributes in the 2nd, 3rd and 4th levels. Letting $K_{leaf,k}$ denote the knowledge score for the i-th leaf attribute in the bottom level, Eq. (5.1) can be simplified as:

$$K = \sum_{k=1}^{n_{leaf}} W_{global,k} \cdot K_{leaf,k}, \quad (5.2)$$

where $n_{leaf} = 19$ is the number of leaf attributes in the assessment framework of Figure 5.1, $K_{leaf,k}$ is

evaluated based on the guidelines in Appendices A-C of appended paper IV , $W_{global,k}$ is the global weight of the k-th "leaf" attribute with respect to the top level goal and is calculated by:

$$W_{global,k} = \begin{cases} W_i \cdot W_{ij}, & if K_{leaf,k} \text{ is in level 3} \\ W_i \cdot W_{ij} \cdot W_{ijk}, & if K_{leaf,k} \text{ is in level 4'} \end{cases} \quad (5.3)$$

Note that the global weights $W_{global,k}, k = 1,2, \dots, n_{leaf}$ of the leaf attributes sums to one: $\sum_{k=1}^{n_{leaf}} W_{global,k} = 1$.

As shown in Appendices A-C of appended paper IV, $K_{leaf,k}$ is between 1 and 5, with a high value indicating strong knowledge. From Eqs. (5.1) and (5.2), and since the scores of leaf attributes are on between 1and 5, it is obvious that also $K_{BE} \in [1,5]$ and a large value indicates strong knowledge on the corresponding BE.

Given the assessment framework developed in Figure 5.1, the AHP [48] is adopted for evaluating the relative importance (weights) W_i, W_{ij} and W_{ijk} in Eq. (5.3). Please note that since there are no alternatives to be compared in this work, pairwise comparison matrices are only needed for deriving the criteria (attributes) weights. More illustration on AHP method and evaluating the weights of criteria is presented in Chapter 2.

As illustrated in Chapter 3, the PRA model is deconstructed to its constituting elements and then, the number of constituting elements is reduced. In this reduced order PRA model, the most basic element is the "basic event", where a minimal cutset consists of a group of "basic events". On the other hand, a given scenario mathematically consists of a group of minimal cutsets. Finally, a given operation states consist of a group of scenarios. Accordingly, the assessment of the SoK starts with the evaluation of the BEs in the reduced-order model. The SoK of the BEs is denoted by K_{BE} and evaluated as in Eq. (5.4) by a weighted average of the leaf attributes scores. We take the generic q-th BE as an example to illustrate step by step the evaluation of SoK assessment method. For the sake of simplicity, we dropped the q subscripts in the symbols:

$$K_{BE} = \sum_{k=1}^{n_{leaf}} W_{global,k} \cdot K_{leaf,k} \quad (5.4)$$

5.3.2. Aggregation of the SoK

Once the SoKs of the basic events in the reduced-order models are evaluated, they can be aggregated to evaluate the total SoK for the PRA model. Let $K_{BE,l,q}$ represent the SoK of the q-th BE in the l-th reduced-order model. The aggregation of $K_{BE,l,q}$ should consider the difference in the atomic elements' (i.e., BEs, MCs, Scenarios, etc.) contribution to the total risk. Different importance measures can be used to

evaluate the contribution of the basic events. For example, as the reduced-order risk model is constructed by the BEs in the MCSs, the weights of the BEs can be calculated based on Fussell-Vesely importance measures [79]:

$$W_{BE,l,q} = \frac{I_{BE,l,q}}{\sum_{q=1}^{n_{BE,l}} I_{BE,l,q}}, \tag{5.5}$$

where $I_{BE,l,q}$ is the Fussell-Vesely importance measure value of the corresponding q-th BE in the elementary risk model l. Remember that "elementary reduced-order risk model" represents the risk model at a given operation state and a given scenario and composed of MCSs at this operation state and scenario, as illustrated in Chapter 3, Eq. (3.11).

The SoK for the l-th elementary reduced-order risk model, denoted by K_l, is calculated by a weighted average of knowledge scores on its basic events by:

$$K_l = \sum_{q=1}^{n_{BE,l}} W_{BE,l,q} \cdot K_{BE,l,q}, \tag{5.6}$$

The importance of the reduced-order model is evaluated by its contribution to the total risk:

$$W_l = \frac{R_{Red,l}}{\sum_{l=1}^{n_l} R_{Red,l}}, \tag{5.7}$$

where $R_{Red,l}$ is the risk index value of the l-th "elementary reduced-order model" and is calculated by Eq. (3.11) in Chapter 3.

To calculate the total SoK K_{Red} of the reduced-order risk model, the knowledge indexes K_ls of the individual reduced-order risk models are further aggregated by considering their contributions:

$$K_{Red} = \sum_{l=1}^{n_l} W_l \cdot K_l, \tag{5.8}$$

The index K_{Red} is, then, used to represent the SoK of the entire PRA of a specific hazard group: its value is between 1 and 5, with a high value indicating that there is strong knowledge in support of the PRA model and its risk outcomes.

5.4. Application

In this section, we apply the developed framework to a case study of real PRA models for two hazard groups in NPPs (previously illustrated in Chapter 3). The reduced-order models that were constructed for each hazard group in Chapter 3 are adopted. The SoK assessment framework is, then, applied on the BEs and the total SoK is obtained by aggregating the BEs' SoKs. Finally, a comparison is made on the SoKs of the two PRA models to provide some conclusions to relevant RIDM.

5.4.1. Reduced-order model

As illustrated in Sect. 5.3, the assessment needs to be carried out at the level of small risk contributors.

Hence, we adopt the developed reduced-order model of the case study presented in Chapter 3 (detailed description of constructing the reduced order model for the same case study is presented in the, Sect. 3.4.2)

5.4.2. Knowledge assessment of basic events

In this section, we show how to assess the SoK for the BEs in Tables 3.4-3.5. As shown in Eq. (5.4), the SoK of the basic event is evaluated as a weighted average over the SoK of the 19 leaf attributes in Figure 5.1. Hence, the first step of applying the SoK assessment framework is to determine the global weights of the "leaf" attributes. The weights are evaluated using the same procedural steps illustrated Chapter 3. Then, the SoK for the "leaf" attributes, i.e., $K_{leaf,k}$ in Eq. (5.4) is determined following the assessment guidelines in in Appendices A-C of appended paper IV. Here, we give an illustrating example on how to evaluate the SoK of the basic event BE2. The first leaf attribute, i.e., quality of assumptions K_{11}, is evaluated based on the guidelines in Appendix A.1 of appended paper IV. In this basic event, the loss of equipment is calculated by assuming that as long as the water reaches the bottom of each equipment, a failure is caused. This assumption is based on extrapolating some data to extreme values, and it is conservative. Therefore, this assumption was judged by the experts to lie between two cases with score 1 and score 3 in Table A.1: an inter-level score of 2 was given by the experts. Take the amount of data K_{21} as another example: the number of years of experience on BE2 is 10 years; therefore, from Appendix B.1 of appended paper IV, the SoK score of K_{21} is assessed by the experts to be 1. The rest of the leaf attributes are assessed similarly and the results are given in Table 5.3 and Table 5.4. Then, from Eq. (5.4) we found $K_{BE} = 3.5500$ for BE2. The procedures are repeated for each BE; the resulting K_{BE}s are given in Table 5.5.

Table 5.3 Assessment of level-3 knowledge "leaf" attributes (BE_2)

Attribute	QA	AD	YE	NE	AE	IN
$W_{i,global}$	0.3234	0.0587	0.1190	0.0630	0.1190	0.1190
Score	2	1	5	5	5	5

Table 5.4 Assessment of level-4 knowledge "leaf" attributes (BE_2)

Attribute	PK	SI	U	RI	PE	PM	P	C	Co	V	Cu	Ac
$W_{global,k}$	0.0203	0.0134	0.0177	0.0144	0.0179	0.0186	0.0221	0.0148	0.0110	0.0147	0.0139	0.0190
Score	5	5	4	4	5	5	4	5	5	3	4	3

5.4.3. Knowledge Aggregation

Finally, the K_{BE}s in Table 5.5 are aggregated for the SoK of the entire model. For this, the SoK of the individual reduced-order risk models K_l need to be calculated first by Eqs. (5.5) and (5.6), with the Fussell-Vesely (FV) importance measures for the BEs also given in Table 5.5. In this case study, we have $l = 1$ for the external events. The resulted K_l from Eqs. (5.5) and (5.6) is $K_l = 2.90$. Then, the total SoK for external flooding, denoted by $K_{Red,Ex}$, is calculated based on the reduced-order model using Eqs. (5.7) and (5.8). In this case study, since we have only one individual risk model, using Eqs. (5.7) and (5.8) leads to $K_{Red,Ex} = K_{l,1} = 2.90$.

Table 5.5 Knowledge assessment and aggregation over the basic events

BE	BE1	BE2	BE3	BE4	BE5	BE6	BE7	BE8	BE9	BE10
FV	0.9020	1.0000	0.5530	0.1820	0.1410	0.1270	0.1210	0.0450	0.0277	0.0277
$W_{BE,l,q} = NFV$	0.2885	0.3199	0.1769	0.0582	0.0451	0.0406	0.0387	0.0144	0.0089	0.0089
K_{BE}	1.6582	3.6595	2.9006	3.2178	3.7778	3.7778	3.0102	3.7778	3.2178	3.2178
$W_{BE,l,q} \times K_{BE,l,q}$	0.4784	1.1705	0.5131	0.1873	0.1704	0.1535	0.1165	0.05437	0.0285	0.0285

*(FV): Fussell-Vesely

*(NFV): Normalized Fussell-Vesely

5.5. Results and discussion

The same steps were repeated on the internal events PRA model. We directly present the final SoK for the internal events PRA model: $K_{Red,In} = 4.04$. The SoK for both hazard groups are graphically illustrated in Figure 5.2. In Figure 5.2, we also illustrate the risk indexes (probability of core meltdown) evaluated for the two hazard groups (note that the values of the risk indexes are scaled due to confidentiality reasons). It can be seen from the Figure 5.2 that the SoK on the internal events is higher than that on external flooding: this means that we are surer of the risk index value calculated with the PRA model of internal events, than of that for the external flooding hazard group.

In fact, these results confirm expectations, as the internal events hazard group has been well studied in nuclear PRAs and mature models are available, whose parameters have relatively low uncertainty [19]. On the other hand, the PRAs for external flooding is generally considered less mature [36] and several limitations have been pointed out in the current external flooding PRA models. For example, the flood frequencies are obtained by extrapolating the fitted historical data (usually limited) to the design basis flood levels, which results in high uncertainty [36]. In particular, the probability of extreme floods is very low [83] and flooding events are very site-specific [86]. Hence, very few data are available for risk

modeling, which limits the SoK for external flooding. The low occurrence probability of external flooding and the lack of operating experience and data related to them makes it very difficult also to predict and estimate their consequences, which adds to the uncertainties in the risk analysis as it limits the SoK of the PRA model used [83]. Specifically, in the case study considered, a large fraction of the risk contribution (69% of the reduced-order risk for external flooding) is due to three basic events i.e., BE_1, BE_2, and BE_3. As shown in Table 5.5, two of them (BE_1, BE_3) have quite low SoK, which limits the SoK of the entire PRA model.

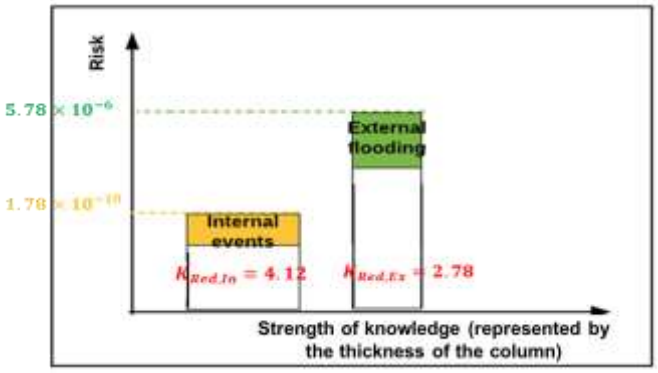


Figure 5.2 Representation of hazard groups' levels of risk and SoK

5.6. Conclusion

In this chapter, we have proposed a new method for implementing a quantitative evaluation of the SoK of risk assessment models. The underlying conceptual framework has been developed based on a thorough literature review. The framework is based on three main attributes (assumptions, data, and phenomenological understanding), which are further decomposed into more tangible sub-attributes and "leaf" attributes for quantification. Detailed scoring guidelines are defined for the evaluation of the leaf attributes. In order to facilitate the application of the knowledge evaluation framework in practice, a top-down bottom-up approach is proposed, where a reduced-order model is constructed in the top-down phase to reduce the complexity of the analysis, and the SoKs are evaluated and aggregated hierarchically in the bottom-up phase. The application of the framework on a real case study of PRA models for two hazard groups, i.e., external flooding and internal events in NPP, has shown its operability. The results of the case study are consistent with the expectations of industrial practice, where the SoK of external flooding is lower than that of internal events, for which more data and information (i.e., strong knowledge) are available.

A potential limitation of the developed method is that we are assuming that the risk assessment model

itself is complete in covering all the possible scenarios. The SoK on model structure and model uncertainty [27], [87] is not considered in this work. For a more comprehensive knowledge assessment, further studies are needed to extend the developed method to consider completeness and comprehensiveness, including model uncertainty in the PRA model [27], [87]. Also, as the weights of the attributes in the framework are subjectively evaluated, formal expert judgment elicitation methods should be used for evaluating the weights. Finally, the evaluation framework and method do not pretend to be complete but they stand as a starting point for a practical assessment of the SoK of risk assessment models.

Chapter 6 Framework for multi-hazards risk aggregation considering the trustworthiness

A criticism of the current practice of MHRA is that the aggregation is conducted by a simple arithmetic summation of the risk metrics from different hazard groups, without considering the heterogeneity in the degrees of maturity and realism of the risk analysis for each hazard group [19]. The risk aggregation should also consider the different realism and trustworthiness in the analyses. In this chapter, we extend the framework developed in Chapter 2 to a more comprehensive and complete framework for trustworthiness assessment. Then, we develop a new method for MHRA considering the level of trustworthiness. In particular, a review of the approaches proposed in the literature for a broader characterization of risk is presented in Sect. 6.1. In Sect. 6.2, a hierarchical framework is developed for assessing the trustworthiness of PRA models. In Sect. 6.3, the procedural steps for implementing the framework are presented. Sect. 6.4 illustrates how to evaluate the risk considering the level of trustworthiness. In Sect. 6.5, the developed framework is applied to a case study from the nuclear industry and finally, Sect. 6.6 concludes this chapter.

6.1. State of the art

It was realized among experts in the domain that a comprehensive representation of the risk is needed to better inform DM. As has been illustrated in Chapter 1, some proposals are found in the literature as an attempt of a broader representation of risk through what so-called “new risk perspectives” that highlights uncertainty instead of probability for representing the risk. We summarize these proposal in the following.

In [8], a structure is presented to help understand the suitability of risk representation through linking the elements of Data-Information-Knowledge-Wisdom hierarchy to the general risk perspectives i.e., events, consequences, uncertainty (A, C, U). In [6], a method is also proposed in accord with the new risk perspective that requires a comprehensive description of risk that covers: the events, consequences, predictions, uncertainty, probability, sensitivity and knowledge. As illustrated in Chapter 4, some attempts are found in the literature for treating uncertain assumptions as an implication of new risk perspectives such as, the law of total expectation, interval probability, crude strength of knowledge and sensitivity

categorization assumption deviation risk [11], [17], [26].

For assessing directly the trustworthiness, we list some of the methods illustrated in Chapter 1. Other methods and detailed description can be also found in Chapter 1. A hierarchical framework is proposed in [29] for evaluating the trustworthiness of risk assessment models through evaluating attributes and sub-attributes of the modeling fidelity and the SoK. In [31], the CMM is proposed to assesses the maturity of a software development process in the light of its quality, reliability, and trustworthiness. A hierarchical framework is proposed in [9] for assessing the maturity and prediction capability of a prognostic method for maintenance DM purposes. A framework for assessing the credibility of M&S is proposed in [9] given eight criteria: (i) verification; (ii) validation; (iii) input pedigree; (iv) results uncertainty (v) results robustness; (vi) use history; (vii) M&S management; (viii) people qualification [12]. The quality of M&S is assessed in ASME by two steps, i.e., verification and validation [32]. Nevertheless, as illustrated previously in Chapter 1, most of the aforementioned works treat the contributing factors to trustworthiness in risk analysis separately, without integrating them in a comprehensive framework that covers all the contributing factors to trustworthiness and they the evaluation of their attributes is carried out by directly scoring the some intangible contributing factors, which is hard to apply in practice. Above all, none of the aforementioned methods integrate the trustworthiness in the result of risk assessment, neither is it considered in MHRA.

6.2. A hierarchical framework for trustworthiness assessment

As illustrated previously, various factors might affect the trustworthiness of risk assessment. We are listing some of the most relevant factors that are believed to greatly affect the trustworthiness of risk assessment. For example, the level of strength of knowledge [6], [8], [17], [7], conservatism [58], [30], uncertainty, level of sophistication and details in the analysis [36], [19], [13], experience, number of approximations and assumptions made in the analysis are identified in [36], [19], [22], [10], [11], [23] as fundamental factors that influence the realism and trustworthiness of analysis.. The communication of the sensitivity is stressed for a comprehensive description of risk [6], [30]. Also, other factors are identified as contributing factors of the credibility of M&S including verification, validation, input pedigree, result's uncertainty, result's robustness, use history, M&S management, people qualification [12].

The trustworthiness of risk assessment is defined in this chapter as the degree of confidence that the background knowledge is strong enough to support the PRA and that PRA model is suitable and correctly made in a robust and thorough way to make the best use of the available knowledge in order to reflect, to

the best possible reality. In this work, a hierarchical tree is developed based on four main factors: (i) the SoK that supports a risk assessment [6], [8], [17], [7]; (ii) the technical adequacy, maturity, quality, and ability of the used tool to represent reality [31], [12], [9]; (iii) the quality of the modeling process [1], [31], [32], [12], [9]; (iv) the sensitivity of the model given the input parameters and assumptions i.e., namely the robustness of the results [6]. The four main factor are categorized into two main groups: the SoK and the modeling fidelity, and in turn broken down more tangible sub-attributes based on a thorough literature review and the attempts presented in the previous chapters. The developed hierarchical framework is presented in Figure 6.1, and detailed definitions of the attributes, sub-attributes and "leaf" attributes are given in Table 6.1-6.4. More information on the attributes elicitation and framework construction are presented in the appended paper V.

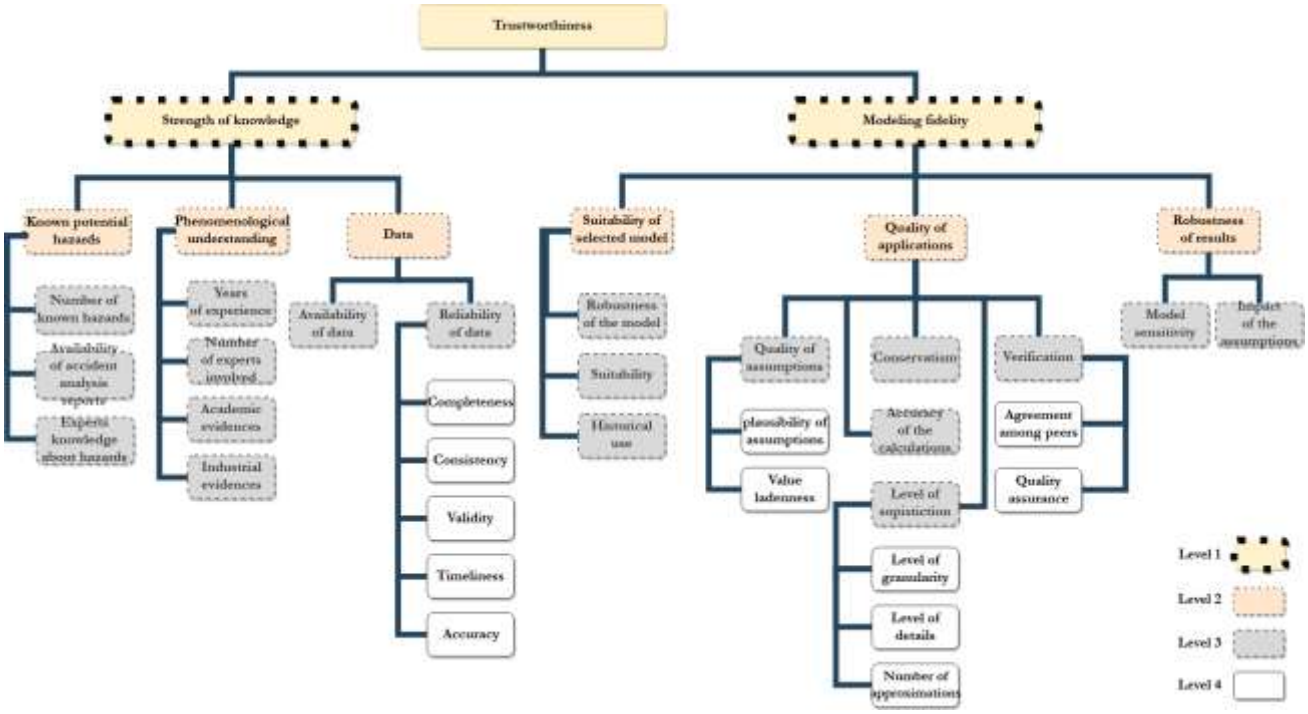


Figure 6.1 A Hierarchical tree for trustworthiness evaluation

Table 6.1 Definition of trustworthiness attributes (Level 1)

Attribute	**Definition**
Modeling fidelity $(MF = T_1)$	The degree of confidence that the selected PRA model is technically adequate for describing the problem of interest and that the model is implemented in a trustable way so that the results of the developed model can reasonably of represents the reality
The strength of knowledge $(SoK = T_2)$	The amount of high-quality explicit knowledge that is available to support the PRA

Table 6.2 Definition of trustworthiness attributes (Level 2)

Attribute	**Definition**
Robustness of the results ($RoR = T_{1,1}$)	The capability of the PRA results to remain unaffected by small variations in model parameters or model assumptions
Suitability of the model ($SoM = T_{1,2}$)	The technical adequacy of the tool, maturity and ability to model the problem of interest
Quality of application ($QAp = T_{1,3}$)	The degree to which the analysis is implemented with the minimum required levels of details and modeling adequacy that have the degree of quality, suitable for supporting the application of interest
Knowledge of potential hazards and accidents evolution process ($PoH = T_{2,1}$)	The availability of documentation and knowledge of abnormal events, accidents and their evolutions, from similar systems
Phenomenological understanding ($Ph = T_{2,2}$)	The knowledge that supports the comprehension of the system functionality and the related phenomena
Data ($D = T_{2,3}$)	Amount and quality of data needed that supports estimating the model parameters

Table 6.3 Definition of trustworthiness attributes (Level 3)

Attribute	**Definition**
Model sensitivity ($MS = T_{1,1,1}$)	The degree to which the model output varies when one or several parameters change
Impact of assumptions ($IoA = T_{1,1,2}$)	The degree to which the model output varies when one or several assumptions change
Robustness of the model ($RoM = T_{1,2,1}$)	The capability of the model to keep its performance when applied to a different problem settings
Suitability of the tool for the problem ($S = T_{1,2,2}$)	The ability to capture all the important details and characterizations of the problem of interest
Historical use ($HU = T_{1,2,3}$)	The degree of confidence gained in this method by the long historical usage
Conservatism ($Cv = T_{1,3,1}$)	The intentional acts for overestimating the risk by making conservative assumptions out of cautiousness
The accuracy of calculations ($AcC = T_{1,3,2}$)	The degree of the voluntarily accepted error in the calculation, e.g., significant figures, simulation errors, and cutoff errors
Quality of assumptions ($QoA = T_{1,3,3}$)	The degree to which the assumption is valid, representing reality and supporting the model
Verification ($Vr = T_{1,3,4}$)	The degree of assurance that the analysis maintains the requirements of quality control standards and obtains the acceptance from different analysts
Level of sophistication ($LoS = T_{1,3,5}$)	The degree of treatment of the problem, and amount of effort and details invested in the problem given its requirement (requirement and complexity)
Number of known hazards ($NH = T_{2,1,1}$)	The documented experience on known hazards that might affect the system of interest
Availability of accident analysis reports ($NH = T_{2,1,2}$)	The availability of technical reports that cover thoroughly the different sequences of any abnormal activity, incident or accident in the time frame and the progressions of each phase
Experts knowledge about the hazard ($NH = T_{2,1,3}$)	The undocumented experience possessed by experts on known hazards

Years of experience ($YE = T_{2,2,1}$)	The amount of experience (measured in years) regarding a specific phenomenon
Number of experts involved ($NE = T_{2,2,2}$)	The number of experts who are explicitly or implicitly involved in understanding the phenomena and the risk analysis
Academic studies on the phenomena ($AE = T_{2,2,3}$)	The number of academic resources, i.e., articles, books, etc., available about the phenomena of interest
Industrial evidence and applications on the phenomena ($IE = T_{2,2,4}$)	The number of industrial applications and reports related to the specific phenomena or events of interest
Amount of available data ($AD = T_{2,3,1}$)	The amount of data that are needed to evaluate the model parameters
Reliability of data ($RD = T_{2,3,2}$)	The degree to which the properties of data satisfy the requirements of risk analysis

Table 6.4 Definition of trustworthiness attributes (Level 4)

Attribute	**Definition**
The plausibility of assumptions ($Pl = T_{1,3,3,1}$)	The degree of realism of the statements made in the analysis, in cases of lack of knowledge or to facilitate the problem solution
Value ladenness of assessors ($VL = T_{1,3,3,2}$)	The experts' degree of objectivity, professionalism, skills and competencies, past fulfillment of assigned missions and level of achievement
Agreement among peers ($Ag = T_{1,3,4,1}$)	The degree of resemblance between the peers on the analysis and assumptions made if they were asked to perform the analysis separately
Quality assurance ($QA = T_{1,3,4,2}$)	The degree of following the standards in the process of implementing the analysis
Level of granularity ($LoG = T_{1,3,5,1}$)	The depth of analysis and subdivision of the problem constituting elements
Number of approximations ($NoA = T_{1,3,5,2}$)	The intentional simplifications made to facilitate the modeling
Level of details ($LoD = T_{1,3,5,3}$)	The degree with which the important contributing factors are captured in the modeling compared to the requirement of the analysis (e.g., the dependency among components)
Completeness ($LoD = T_{2,3,2,1}$)	The degree to which the collected data contain the needed information for the risk modeling and assessment
Consistency ($LoD = T_{2,3,2,2}$)	The degree of homogeneity of data from different data sources
Validity ($LoD = T_{2,3,2,3}$)	The degree to which the data are collected from a standard collection process and satisfy the syntax of its definition (documentation related)
Timeliness ($LoD = T_{2,3,2,4}$)	The degree to which data correctly reflect the reality of an object or event
Accuracy ($LoD = T_{2,3,2,5}$)	The degree to which data are up-to-date and represent reality for the required point in time

6.3. Evaluation of the level of trustworthiness

In this section, a bottom-up method for evaluating the level of trustworthiness is developed where a combination of Dempster Shafer Theory (DST) and Analytical Hierarchy Process (AHP) is used to determine the weights of the attributes/sub-attributes in the framework proposed in Figure 6.1.

6.3.1. Evaluation of the trustworthiness

In this framework, five levels of trustworthiness are defined with their corresponding settings:

1. 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 DM.

2. 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 DM.

3. 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.

4. 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.

5. 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 DM 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.

In this work, the level of trustworthiness is calculated using a weighted average of the “leaf” attributes in

Figure 6.1.

$$T = \sum_i^n W_i \cdot A_i \tag{6.1}$$

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 Appendices of the appended paper V; n is the number of the leaf attributes (in Figure 6.1, we have $n = 27$). The scores should be assigned using the scoring guidelines presented in Appendices A-B of the appended paper V. On the other hand, the weights are determined based on Dempsthe Shafer-Analytical Hierarchy Process (DST-AHP) as discussed in Sect. 6.3.2 [88].

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

The weights of the different attributes in Figure 6.1 can be determined by using the AHP method to compare their relative importance with respect to the trustworthiness of risk assessment [48]. AHP is usually 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, 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 caused by the multidisciplinary nature of the problem or 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 [88], [89]. In this method, the assessors are asked to identify the focal sets that comprise of a single or group of the 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 DST fusion rule. The weights for each criterion are assumed to be BBA of the corresponding focal element (single criterion), and are derived based on maximum belief-plausibility principle in Dempster-Shafer theory, or on the maximum subjective probability obtained by probabilistic transformations using the transferable belief model [88], [90], [89]. It should be noted that in this work, we

only apply this method to derive the relative weights of the criteria, rather than using it to rank alternatives. Similar ideas have been used in [91], [92]. Procedure for calculating the weights of the leaf attributes based on DST-AHP is presented below.

I. 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 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 three modeling fidelity daughter attributes:

$$\begin{array}{c} \\ T_{1,1} \\ T_{1,2} \\ T_{1,3} \end{array} \begin{array}{c} \begin{array}{ccc} T_{1,1} & T_{1,2} & T_{1,3} \end{array} \\ \begin{bmatrix} 1 & MF_{12} & MF_{13} \\ MF_{21} & 1 & MF_{23} \\ MF_{31} & MF_{32} & 1 \end{bmatrix} \end{array}$$

where the entries 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 generic element MF_{ij} is assigned by assessing the relative importance of attribute i to attribute j following the scoring protocols in [48]. 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 [93], [92], [89]. 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 6.1-6.4). 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 [46].

II. Computing the pairwise comparison matrix

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 [94]. 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.

III. 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 [89]. Shafer's reliability discounting is often used to consider the reliability of the source information in DST-AHP [95]:

$$m_\delta(A) = \begin{cases} (\delta) \cdot m(A) & \forall A \subseteq \Theta,\ \mathrm{A} \neq \Theta \\ (1-\delta) + (\delta) \cdot m(\Theta), & \mathrm{A} = \Theta \end{cases},\ \delta \in [0,1] \tag{6.2}$$

where Θ represents the complete set of criteria, A is the focal elements 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.

IV. Combination of experts opinions

Next, Dempster's rule of combination [95] is used to combine two independent pieces of evidence assigned by different experts. The discounted BBAs from different experts are combined by [89]:

$$m_{1,2}^\delta(C) = \left(m_1^\delta \oplus m_2^\delta\right)(C) = \begin{cases} 0 & C = \phi, \\ \frac{1}{1-K} \cdot \sum_{A \cap B = C \neq \phi} m_1^\delta(A) \cdot m_2^\delta(B) & C \neq \phi, \end{cases} \tag{6.3}$$

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:

$$K = \sum_{A \cap B = \phi} m_1^\delta(A) \cdot m_2^\delta(B) \tag{6.4}$$

V. 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 paper, the transferable belief model proposed by [96] 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 [96], [97]:

$$w(x) = \sum_{C \subseteq \Theta, C \neq \phi} m(C) \frac{1_{C(x)}}{|C|}, \forall x \in \Theta \tag{6.5}$$

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:

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

where $w(HU), w(SoM)\ and\ w(MF)$ are the local weights of the historical use, the suitability of model, and the modeling fidelity. For simplicity reasons, hereafter the global weights for leaf attributes are denoted by W_i and in the framework of Figure 6.1, we have $i = 1,2,\cdots,27$.

6.4. Evaluation of the risk considering trustworthiness levels

In this section, the "weighted posterior" method is used for integrating the risk index with the trustworthiness of the PRA for a single hazard group and a structured methodology is developed for eliciting these weights. Finally, an illustration is presented on MHRA considering the level of trustworthiness.

6.4.1. Evaluation of the risk of a single hazard group

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 [98]. 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:

$$Risk|T = P(E_T) \cdot \text{Risk}|E_T + (1 - P(E_T) \cdot \text{Risk}|E_{NT} \quad (6.6)$$

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; $\text{Risk}|E_T$ is the estimated risk from the PRA. Due to the presence of epistemic (parametric) uncertainty in the analysis, $\text{Risk}|E_T$ is often expressed as a subjective probability distribution of the risk index. $\text{Risk}|E_{NT}$ is an alternate distribution of the risk when the decision maker thinks the PRA is not trustable. In this paper, we assume $\text{Risk}|E_{NT}$ 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 [99]. For example, [100] developed a similar model to calculate the default risk in similar scenarios considering the unexpected events.

The following steps summarize how to use Eq. (6.6) to evaluate the risk given the trustworthiness of the risk assessment:

i. The risk distribution $\text{Risk}|E_T$ is evaluated for each hazard group using conventional PRA considering the parametric uncertainty propagation.
ii. The level of trustworthiness of PRA of the corresponding hazard group is assessed, using the procedures in Section 6.3.
iii. The subjective probability of trusting the PRA is determined by the detailed procedures described in Section 6.4.2.
iv. The level of trustworthiness is integrated in the risk using Eq. (6.6).

6.4.2. Determining the probability of trusting the PRA

The probability $P(E_T)$ in Eq. (6.6), 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. [101]. In this work, we choose the "certainty equivalent gambles" for the elicitation. We summarize the following steps for the elicitation of the probability of trust using the "certainty equivalent gambles" and some general recommendations are

presented in the appended paper V for ensuring the quality of the elicitation process:

i. The elicitor informs the decision maker about the definition of the different levels of trustworthiness and its physical meaning, based on the definitions in Sect. 6.3.1.

ii. The decision maker is asked to compare two scenarios: (1) he/she participates in a gamble where he/she will win $1,000 if an accident occurs and $0 if the accident does not occur; (2) he/she wins $x for sure.

iii. The experts exchange information between them and discuss.

iv. Suppose that a PRA was conducted and predicts that the consequences will occur for sure, and the trustworthiness of the PRA is one of the five levels defined in Sect. 6.3.1. Then, for each level of trustworthiness, the elicitor varies the value of x until the decision maker feels indifferent between the two scenarios.

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

$$p = \frac{x}{1000} \tag{6.7}$$

where 1000 here represents the $1000 that the expert gains if the accident does not occur (the model prediction is correct).

vi. 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.

The risk assessor can eventually use this function to estimate the probabilities of trust for each hazard group.

6.4.3. MHRA considering trustworthiness levels

Main steps for MHRA considering the level of trustworthiness are presented in Figure 6.2.

Trustworthiness of each single group PRA is evaluated and integrated into the risk estimate for each 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 simply done by adding the risk distributions from different hazard groups, as shown in Eq. (6.8), where $Risk_{total}$ is the total risk considering the level of trustworthiness; $(R_i|\mathrm{T})$ is the risk from the hazard group i given the level of trustworthiness; n is the number of hazard groups. Monte-Carlo simulations are often used to do the summation.

$$Risk_{total} = \sum_{i=1}^{n}(Risk_i|\mathrm{T}) \tag{6.8}$$

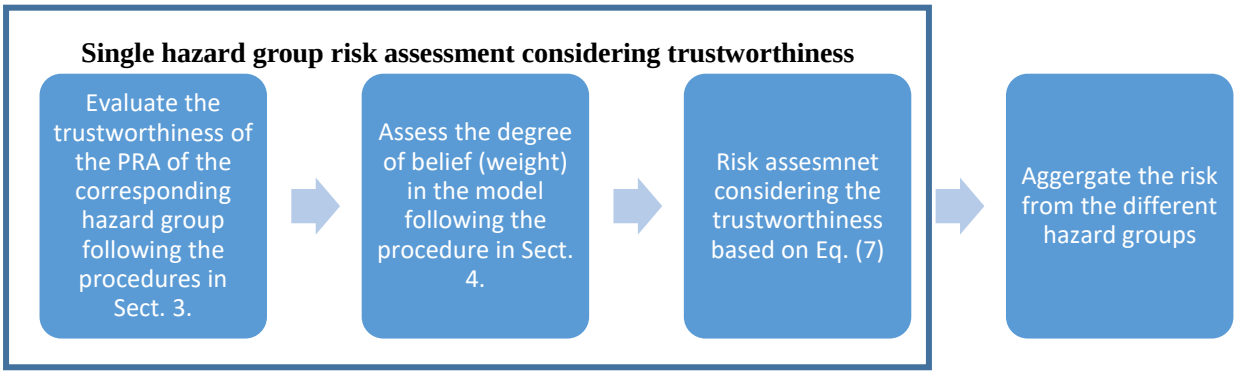


Figure 6.2 Main steps for MHRA considering the trustworthiness of the PRA

6.5. Application

In this section, we apply the developed framework to 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 group were developed and provided by EDF. The level of trustworthiness was then, assessed for each hazard group. The risk distributions from each hazard group were then recalculated considering the level of trustworthiness, and finally, the risk was aggregated from the two hazard groups.

6.5.1. Description of the PRA model

The two hazard groups considered in this framework are external flooding and internal events. The external flooding refers to the overflow of water that is caused by naturally induced hazards such as river overflows, tsunamis, dam failures and snow melts [83], [102]. The internal events refer to any undesired event that originates within the NPP and can cause initiating events that might lead to abnormal state and eventually, a core meltdown [19]. Examples of internal events include structural failures, safety systems operation and maintenance errors, etc. [84]. In this case study, bow-ties models are used to assess the probability of Core Damage Frequency (CDF). In this case study, the risk analysis was provided by EDF [7]. In the original work of EDF, the uncertainty propagation was implemented, but only the mean values

of the probability distributions of the risk were considered in MHRA and used for comparison to the safety criteria. However, due to confidentiality reasons, real values cannot be presented. Instead, we artificialize the risk distribution for illustration purposes. The risk distributions with parametric uncertainty propagation are presented in Figure 6.3.

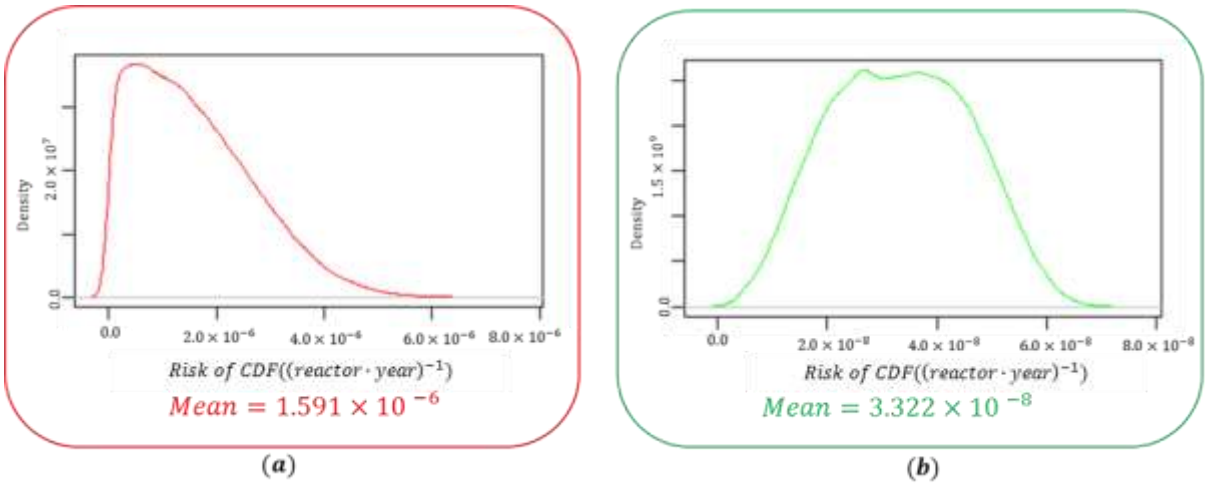


Figure 6.3 Probability distribution of the risk considering the parametric uncertainty: (a) external flooding risk, (b) internal events

6.5.2. Evaluation of level of trustworthiness

6.5.2.1. Evaluation of the attributes weights

As illustrated in Sect. 6.3, the first step for evaluating the level of trustworthiness is to determine the relative importance (weights) of the trustworthiness attributes. The weights of the attribute are evaluated using DST-AHP technique. Here, for illustration reasons, the sub-attribute "modeling fidelity" (T_1) is taken as an example to illustrate how to obtain local weights through pairwise comparison and DTS-AHP.

I. Constructing pairwise comparison matrices

As shown in Sect. 6.3, the first step in DST-AHP technique is to construct the pairwise comparison matrix. Take the daughter attributes of modeling fidelity as an example. In this example, a 4×4 pairwise comparison matrix is constructed in Table 6.5.

Table 6.5 Pairwise comparison matrix (knowledge matrix) for comparing modeling fidelity "daughter" attributes

Modeling fidelity	$\{T_{1,1}\}$	$\{T_{1,2}\}$	$\{T_{1,3}\}$	$\Theta = \{T_{1,1}, T_{1,2}, T_{1,3}\}$
$\{T_{1,1}\}$	1	0	0	1/2
$\{T_{1,2}\}$	0	1	0	5/2
$\{T_{1,3}\}$	0	0	1	4

$\{T_{1,1}, T_{1,2}, T_{1,3}\}$	2	2/5	1/4	1

Please note that the zeros that appear in the matrix indicate that there is no need to compare the individual criteria directly: they are compared indirectly through comparing the individual criteria to the universal set Θ [88].

$T_{1,1}$ represents the Quality of application, $T_{1,2}$ represents the Suitability of the model, $T_{1,3}$ represents the robustness of the results

In this matrix, the expert has considered four groups of focal sets: three for individual criteria and one containing all the criteria in order to consider the uncertainty in the evaluation. Choosing focal sets like this means that to the best of their knowledge, the experts believe that the aforementioned focal sets can be favorably compared to the universal set Θ.

II. Computing the pairwise comparison matrix

In the previous example, the expert was asked to fill the pairwise comparison matrix to express his/her preference of a criterion over another. In this step, the weights of the focal sets are derived using conventional AHP technique, where the normalized principal eigenvector of the matrix represents the weights. This can be directly done by normalizing each column in the matrix individually and, then, averaging the elements in each row to obtain that weight.

Table 6.6 Normalized pairwise comparison matrix (knowledge matrix) of modeling fidelity "daughter" attributes

Modeling fidelity	$\{T_{1,1}\}$	$\{T_{1,2}\}$	$\{T_{1,3}\}$	$\Theta = \{T_{1,1}, T_{1,2}, T_{1,3}\}$	Weight (BBA)
$\{T_{1,1}\}$	0.33	0	0	0.06	0.10
$\{T_{1,2}\}$	0	0.71	0	0.31	0.26
$\{T_{1.3}\}$	0	0	0.8	0.5	0.32
$\{T_{1,1}, T_{1,2}, T_{1,3}\}$	0.67	0.29	0.2	0.13	0.32

III. Reliability discounting

After computing the BBA for each expert's matrix, they need to be discounted based on the reliability of each expert. For illustration purposes, the reliability δ of the expert who made the assessment is assumed to be 0.60. From Eq. (6.2), the discounted weights are found as the following:

$$m_{0.60}(T_{1,1}) = 0.6 \times 0.10 = 0.06$$

Similarly for $m_{0.60}(T_{1,2}) = 0.16$, & $m_{0.60}(T_{1,3}) = 0.19$.

Finally, $m_{0.60}(\Theta)$ is found as the following:

$$m_{0.60}(\Theta) = (1 - 0.60) + 0.6 \times 0.32 = 0.59$$

Please note that the BBAs (weights) sum to one before and after the discounting.

IV. Combination of experts opinions

In This case study, three experts are invited for evaluating the weights. Their assigned BBAs are summarized in Table 6.7 (the BBAs are calculated following the steps in Sect 6.3.2, step III).

Table 6.7 discounted basic belief assignment from two experts

Focal sets of the criteria	Expert 1	Expert 2	Expert 3
	$m_\delta(A)$	$m_\delta(A)$	$m_\delta(A)$
$\{T_{1,1}\}$	0.06	0.16	0.02
$\{T_{1,2}\}$	0.16	0.24	0.38
$\{T_{1,3}\}$	0.19	0.24	0.46
$\{T_{1,1}, T_{1,2}, T_{1,3}\}$	0.59	0.36	0.14

The combination of the experts' judgments is conducted sequentially. Table 6.8 shows the procedures for combining the judgments of the first two experts.

Table 6.8 Dempster's rule of combination matrix

Expert 1 \ Expert 2	$m_\delta(T_{1,1})$	$m_\delta(T_{1,2})$	$m_\delta(T_{1,3})$	$m_\delta(T_{1,1}, T_{1,2}, T_{1,3})$
$m_\delta(T_{1,1})$	$m_\delta(T_{1,1})_1$	ϕ_1	ϕ_2	$m_\delta(T_{1,1})_2$
$m_\delta(T_{1,2})$	ϕ_3	$m_\delta(T_{1,2})_1$	ϕ_4	$m_\delta(T_{1,2})_2$
$m_\delta(T_{1,3})$	ϕ_5	ϕ_6	$m_\delta(T_{1,3})_1$	$m_\delta(T_{1,3})_2$
$m_\delta(T_{1,1}, T_{1,2}, T_{1,3})$	$m_\delta(T_{1,1})_3$	$m_\delta(T_{1,3})_3$	$m_\delta(T_{1,3})_3$	$m_\delta(T_{1,1}, T_{1,2}, T_{1,3})$

*Please note that the element ij in the table represent the multiplication of the elements $1j \times i1$, e.g., $m_\delta(T_{1,1}) \times m_\delta(T_{1,1}) = m_\delta(T_{1.1})_1$

From Eq. (6.4), $K = 0{,}17$.

From Eq. (6.3):

$$m_{1,2}^{\delta}(T_{1,3}) = \frac{0{,}26}{1 - 0.17} = 0.31$$

The same steps are repeated for the other mass functions and presented in Table 6.9. Finally, the new results obtained from the combination of the two experts are used to be combined with the BBAs from the third expert matrix. The results are presented in Table 6.9.

Table 6.9 Mass function combinations from the experts

Focal sets of the criteria	Combined mass from experts 1 and 2	Combined mass from experts 1, 2 and 3
	$m_\delta(A)$	
$m_{1,2}^{\delta}(T_{1.1})$	0.31	0.49
$m_{1,2}^{\delta}(T_{1.2})$	0.29	0.40

$m_{1,2}^{\delta}(T_{1.3})$	0.15	0.05
$m_{1,2}^{\delta}(T_{1.1}, T_{1.2}, T_{1.3})$	0.25	0.06

V. Pignistic probability transformation

Then, the pignistic mass function is found by Eq. (6.5):

$$w_{1,2,3}^{\delta}(T_{1,1}) = m_{1,2,3}^{\delta}(T_{1,1}) + \frac{m_{1,2,3}^{\delta}(T_{1,1}, T_{1,2}, T_{1,3})}{3} = 0.05 + \frac{0.06}{3} = 0.07$$

The steps are repeated for the other mass functions and found to be:

$$w_{1,2,3}^{\delta}(T_{1,2}) = 0.42$$

$$w_{1,2,3}^{\delta}(T_{1,3}) = 0.51$$

Note that the three mass functions on the pignistic level sum to one. These pignistic mass functions represent the relative "believed weights" of the three criteria under modeling fidelity after the reliability discounting and transformation. The same steps are repeated for all the criteria. Then, the weights need to be evaluated with respect to the top-level goal: the trustworthiness. As illustrated previously, this can be done easily by multiplying the weight of the daughter attribute by the weight of the upper parent attributes in each level. For simplicity reasons, only the weights of the "leaf" attribute with respect to the top level attribute i.e., trustworthiness, are presented in Table 6.10 and 6.11. Note that the weights of the 27 leaf attributes with respect to the top goal, sum to one $\sum_{i}^{27} W_i = 1$.

6.5.2.2. Evaluation of the attributes scores

The next step for evaluating the level of trustworthiness is to evaluate the attributes score for the hazard group, given the scoring guidelines in Appendices A-B of the appended paper V. Some information regarding the risk assessment process is extracted from the PRA report to support the trustworthiness assessment.

- The heights (water levels) at the plant's platform at which the water can lead to a failure of a specific element were defined.
- The water flowrate that would result in a given water height at the NPP platform in a defined interval of time was predicted.
- The flow-rate was multiplied by a safety factor of 130%.
- The "return period" for each flowrate was obtained from the data of the millennial flooding flowrate of the river of interest, and the data were extrapolated to assess the frequencies of extreme flowrates.
- The river flooding is considered as a predictable phenomenon and the probability of failure of

transition into the emergency state (i.e., normal shutdown and cooling with steam generator, residual heat removal system, etc.) is assumed to be the intrinsic probability of failure.

- It is assumed that river overflow is the only source of external flooding.
- A combined hydraulic/hydrologic method is adopted, given the special hydrological and physical characteristics of the basin.
- It is assumed that once the water reaches the bottom of the equipment, the equipment fails.
- It is assumed that failing to close the valves (ensuring the volumetric protection sealing-water proofing) causes the total loss of Emergency Feedwater System (EFWS).
- It is assumed that clogging inevitably occurs if the flooding occurs.
- The analysis and model calculation for this hazard group is taken with a specific cutoff error of 10^{-14}.

Based on the excerptions from the report, it can be seen that:

- In this example, the risk analysis and assessment steps follow the IAEA recommendations.
- The calculation of flowrates and flow frequencies are calculated using solid deterministic models. However, extrapolation of the data to obtain the frequencies of floods with extreme flowrates is still doubtful.
- The river overflow is a predictable phenomenon and does not happen suddenly. However, the river overflow is not the only source of flooding. For example, a rupture in the river dikes might also lead to sudden, unpredictable flooding.
- The application of a combined hydraulic/hydrologic method on the flooding studies of nuclear sites allows a more realistic evaluation of the flooding level and to estimate more precisely the return periods.
- The assumption that the water will fail the equipment directly if it touches its bottom level is conservative.
- Feedback data show that clogging due to river flooding has occurred before in the nuclear industry (see, for example, USNRC General Electric Advanced Technology Manual for more information [103]). However, claiming that each flooding would surely lead to clogging is still questionable and needs to be studied in details, taking into account the different influencing parameters (hydraulic, geometrical and topographical properties) of the area (see [104]).

- In case of failing to close the valves ensuring the volumetric protection, the probability that water will go back through the drainage system is not identified and assumed to be one ($P = 1$), though there are no relevant calculations. Moreover, once the water enters the physical protection locations, the safety-related equipment is assumed to be lost. Both assumptions are conservative to increase the safety margin.

Based on the above observations, the leaf attributes in Figure 6.1 can be evaluated. For example, quality assurance attribute is evaluated to be five ($T_{1,3,4,2} = 5$), since the PRA is conducted following the IAEA recommendations. The accuracy of the calculation is evaluated to be five ($T_{1,3,2} = 5$), since the cutoff error is apparently very low. The combined hydraulic/hydrologic models used for the flooding studies are able to capture the special hydrological and physical characteristics of the basin, which makes them suitable for the study. Hence, a score of four ($T_{1,2,2} = 4$) is given for the suitability of the model. The assumptions presented above are mostly conservative and unrealistic. Therefore, a score of one ($T_{1,3,3,1} = 1$) is given for the plausibility of the assumptions. The other attributes are scored in the same way. The results are represented in Tables 6.10 and 6.11. The level of trustworthiness for the external flooding is, then, calculated by Eq. (6.1): $T_{ext} = \sum_{i=1}^{27} W_i \cdot A_i = 3.260$.

Table 6.10 level-3 leaf attributes weights W and scores S for external flooding hazard group

Att	*MS*	*IoA*	*RM*	*S*	*HU*	*Cv*	*AoC*	*NH*	*AR*	*EK*	*YE*	*NE*	*Ac*	*In*	*AD*
W	0.01	0.02	0.02	0.15	0.07	0.02	0.01	0.02	0.03	0.05	0.03	0.01	0.10	0.10	0.06
	2	6	5	8	0	5	2	2	2	4	4	7	5	5	5
Score	2	2	3	4	3	4	5	2	2	3	3	4	3	3	3

Table 6.11 level-4 leaf attributes weights W and scores S for external flooding hazard group

Att	*Pl*	*VL*	*Ag*	*QA*	*LoG*	*NoA*	*LoD*	*C*	*Co*	*V*	*T*	*Ac*
W	0.037	0.029	0.025	0.066	0.006	0.005	0.004	0.017	0.011	0.009	0.011	0.017
Score	1	4	4	5	4	4	4	3	3	3	3	3

The trustworthiness for internal events hazard group (T_{int}) was calculated in the same way and, the result is $T_{int} = 4.414$. These results confirm the expectations, where the PRA for internal events is considered relatively mature and well established [19] in contrast to the PRA of external hazards which, is considered less mature with several limitations [36].

6.5.2.3. Determining the probability of trust in the PRA results

In this step, the decision maker is asked to assign a probability that represents their belief that the risk assessment model output is correct, based on the certainty equivalent approach presented in Sect. 6.4.2. The results given by the experts are given in Table 6.12. The data in Table 6.12 are extrapolated and fitted to a function, as shown in Figure 6.4. As illustrated in Figure 6.4, the expert exerts a risk neutral behavior.

Table 6.12 Probability of trust given the level of trustworthiness

Trustworthiness	Probability of trust
1	0.05
2	0.50
3	0.75
4	0.90
5	1.00

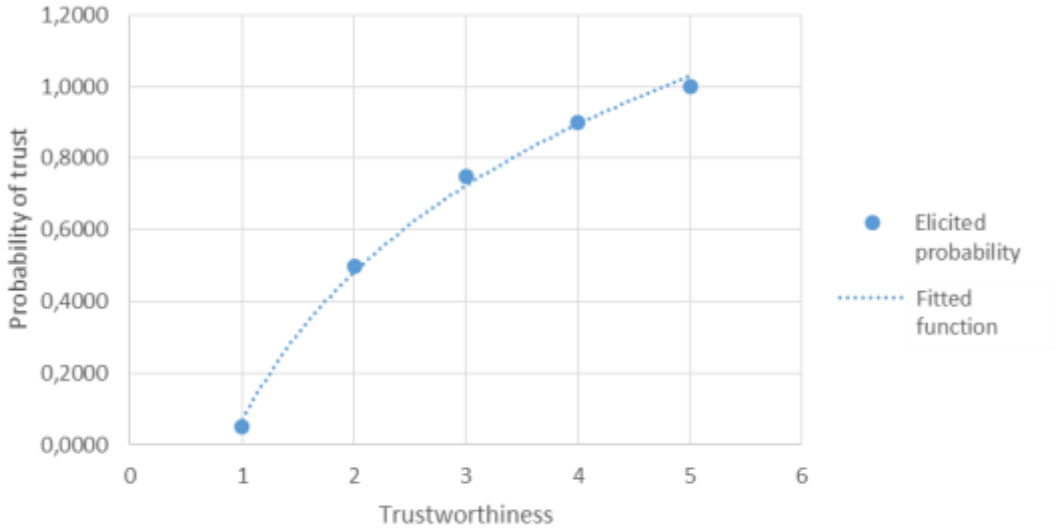


Figure 6.4 Fitted probability of trusting the PRA given the trustworthiness

Then, the probability that the decision maker trusts each hazard group PRA given their trustworthiness is calculated from the fitted model in Figure 6.4. The probability of trust for the external flooding p_{ext} is found to be $p_{ext} = 0.783$. The probability of trust for the internal events p_{int} is found to be $p_{int} = 0.957$.

6.5.2.4. Multi-Hazards risk aggregation the level of trustworthiness

The level of trustworthiness is integrated with the PRA results for both hazard groups following Eq. (6.6). The results are presented in Figure 6.5-6.6, respectively. As can be seen from Figure 6.5 (a), which represents the risk analysis results considering only the parametric uncertainty in the analysis, most of the mass of the risk distribution concentrates in the narrow interval of $[4.626 \times 10^{-11}, 7.738 \times 10^{-6}]$. After

integrating the level of trustworthiness, however, the interval increases to $[3.019 \times 10^{-6}, 2.169 \times 10^{-1}]$ (Figure 6.5 (b)). The mean risk value for external flooding considering the trustworthiness is 1.088×10^{-1} $(reactor \cdot year)^{-1}$, compared to 1.589×10^{-6} $(reactor \cdot year)^{-1}$ without considering it. For internal events, a similar effect is seen in Figure 6.6 (the mean risk value is 2.149×10^{-2} $(reactor \cdot year)^{-1}$ considering the trustworthiness compared to 3.322×10^{-8} $(reactor \cdot year)^{-1}$ without considering it). It is, then, seen that considering the level of trustworthiness leads to a larger spread-out of the probability distribution of the risk.

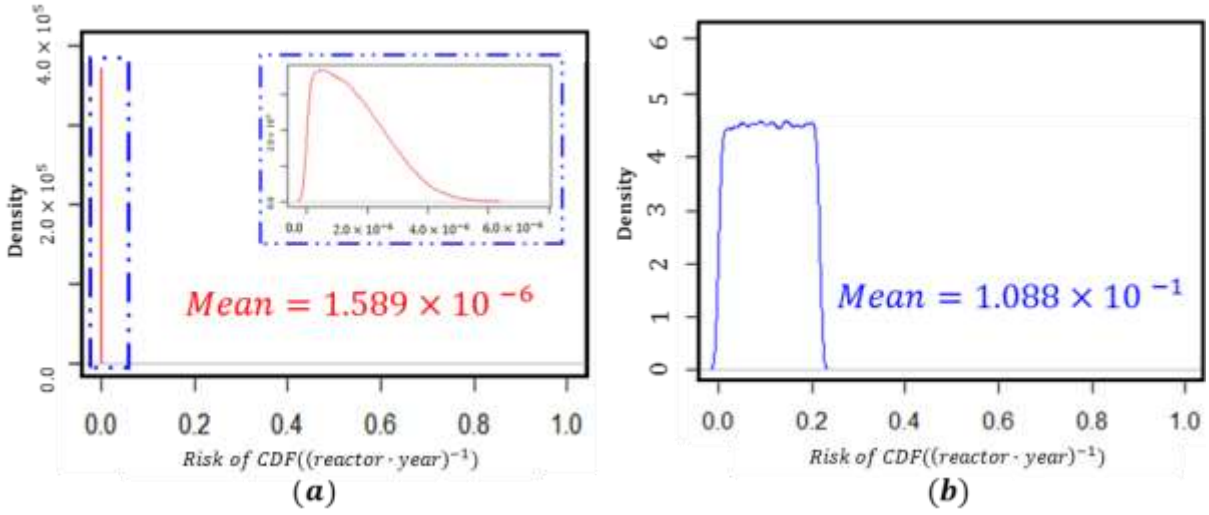


Figure 6.5 Updated risk estimates after considering the level of trustworthiness for external flooding (a) original risk estimate from the PRA, (b) Risk estimates after integrating the level of trustworthiness

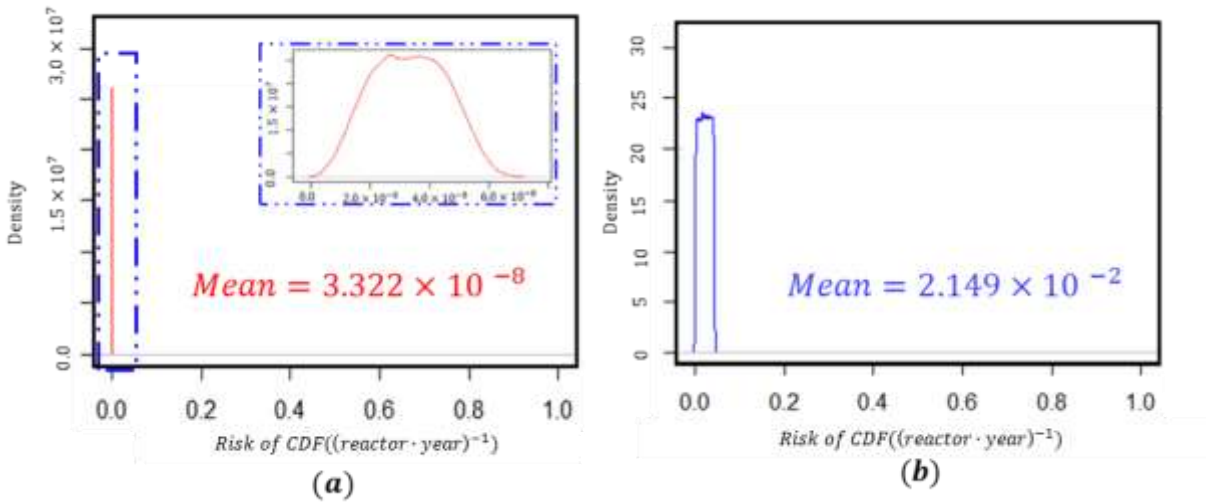


Figure 6.6 Updated risk estimates after considering the level of trustworthiness for internal events (a) original risk estimate from the PRA, (b) Risk estimates after integrating the level of trustworthiness

6.5.2.5. Multi-Hazards risk aggregation

Finally, the overall risk given the level of trustworthiness can be calculated using Eq. (6.8). The results are presented in Figure 6.7. The empirical probability density function of the risk is obtained through a Monte-Carlo simulation of 10^5 samples. The mean value of the total risk from the two hazard groups considering the level of trustworthiness is found to be $1.303 \times 10^{-1}\ (reactor \cdot year)^{-1}$ compared to $1.622 \times 10^{-6}\ (reactor \cdot year)^{-1}$ without considering it. 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. The increase of the spread-out of probability distribution of risk leads to a higher mean value of risk. The aggregation of the risks from the two hazard groups considering the level of trustworthiness results in a more meaningful result as it takes into account the fact that the PRA model of the two hazard groups is based on different levels of trustworthiness.

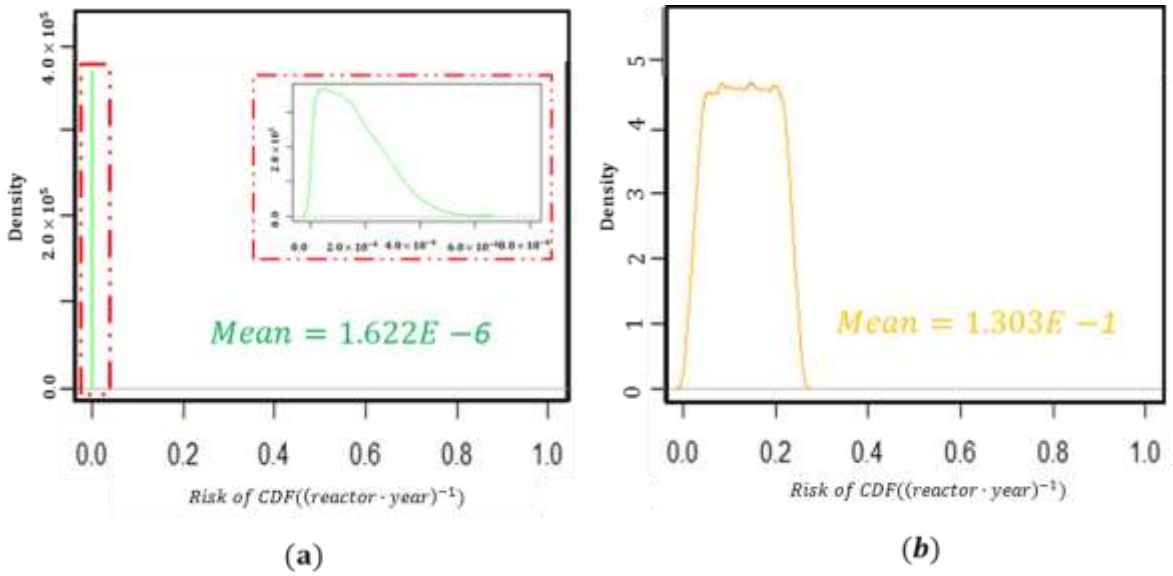


Figure 6.7 Results of the MHRA, (a) conventional aggregation, (b) considering the level of trustworthiness

6.6. Conclusion

In this chapter, we have presented a framework for MHRA considering trustworthiness. A framework for evaluating the level of trustworthiness is first developed. The framework consists of two main attributes, i.e., the strength of knowledge and modeling fidelity. The strength of knowledge attribute covers the explicit knowledge that can be documented, transferred or explained. The modeling fidelity attribute covers the suitability of the tool and the model construction process. The two attributes are broken down into sub-attributes and, finally, leaf attributes. The total trustworthiness is calculated using a weighted average of the attributes, where the weights are calculated using DST-AHP method, in which the AHP

method is used to calculate the relative weights of the attributes using experts' elicitations, whereas the DST method is used to account for the uncertainty in the elicitation.

A MHRA method is, then, developed to aggregate the risk from different hazard groups with different levels of trustworthiness, based on a "weighted posterior" method. An application to a case study of a NPP shows that the developed method allows aggregating risk estimates with different degrees of maturity and realism from different risk contributors.

Chapter 7 Conclusion and future work

7.1. Conclusion

The objective of risk assessment is to provide informative support to DM [35], [36], [5], [3], [34]. In risk assessment, we perform quantitative and qualitative measures of risk to ensure that it is maintained under the allowed safety limit. The quantitative evaluation of risk is done by MHRA, which includes aggregating the risk indexes from different contributors to arrive at a risk metric that can be compared to the safety criteria to support DM. On one hand, in MHRA, the risk indexes from different contributors might have different degrees of realism, which, in turn, results from differences in characterizations, e.g., of uncertainty, background knowledge, conservatism, etc. [19]. On the other hand, the current practice of MHRA consists of a simple arithmetic summation of the risk indexes from the different contributors without considering the aspects that lead to the difference in the degrees of realism [19]. MHRA must, therefore, consider their different uncertainties [19] and the confidence on the outcomes that is relevant to support DM [3].

In this thesis, we focus on enhancing the description and evaluation of risk for a more assured practice of RIDM. In particular, we have provided a methodological framework for MHRA and the assessment of the level of trustworthiness, which a risk assessment is based upon. The following specific contributions have been attained:

1. Important factors contributing to the trustworthiness of risk assessment have been identified;
2. An integrated hierarchical framework has been developed for systematically organizing these factor for the assessment of the trustworthiness of risk outcomes;
3. A technique based on DST-AHP has been adapted to consider the assessors subjectivity in the assessment process;
4. A MHRA technique based on Bayesian model averaging has been developed to integrate the trustworthiness of individual hazard groups' risk outcomes for informed decision making.

The developed framework provides a systematic way to evaluate the trustworthiness in risk assessment outcomes and integrate it in the results of risk aggregation to overcome the shortcomings of conventional MHRA. From a practical point of view, the framework also provides systematic and practical procedures that facilitate the application to real cases and overcomes the problem of subjectivity in experts' judgments. The application of the developed framework to real life case studies demonstrates the

feasibility and reasonableness of the approach, paving the way for its potential applicability to inform risk-based decision making.

7.2. Discussion

The framework in Chapter 6 was developed by considering different models to evaluate the factors relevant to trustworthiness (Chapters 2-5). Detailed definitions of the attributes in the hierarchical framework have been introduced and their assessment have been illustrated (see for example Table 2.1 and Tables 6.1-6.4) to check feasibility in practice.

The attributes have been elicited in two ways in an effort to ensure completeness: deductive and inductive reasoning. The deductive reasoning was based on literature survey, expert elicitation and deep reasoning: a large number of candidate attributes are collected and, then, screened based on their relevance to trustworthiness. The inductive reasoning was based on deducting the elements needed to construct the risk assessment model. The most important and representative attributes have, then, been studied individually to understand their effect on trustworthiness and to study the possibility of a more granular and comprehensive evaluation that covers all possible sub-attributes (Chapters 4-5).

The relative importance (weights) and scores of the attributes in the framework are assessed based on experts' elicitation. Several factors affect the consistency and quality of experts' judgments, e.g., lack of prior knowledge on the problem, subjectivity of judgments and delicacy of the subject, and the fact that experts make judgments not only on the criteria of their specialty, but also about all other criteria [105]. To ensure the quality and consistency of experts' judgments, a rigorous evaluation procedure has been introduced along with predefined evaluation protocols. The procedural steps introduced allow improving the quality of the information provided to select the experts needed to make the judgments, as well as the quality of information required to assess the attributes. In addition, a behavioral and a mathematical aggregation technique has been introduced to consider the uncertainty in the experts' judgments and enhance the quality and consistency in their judgments (Chapter 6). The evaluation protocols were established based on technical reports (Chapters 2-6), literature, and experts' knowledge, so that the consistency of the evaluation can be ensured to the maximal degree. Although the subjectivity in the evaluation cannot be eliminated, the developed methodology is an attempt to enhance its consistency and quality through a systematically organized evaluation process.

7.3. Future work

The framework presented in the thesis has been shown feasible through the application to real case

studies. However, there are still issues that need to be worked out. For example, the assessment of some factors was conducted semi-quantitatively, using evaluation guidelines, but remaining subjective at large. Efforts should be devoted in enhancing the assessment guidelines and developing rigorous enumerating (assessment) protocols to further reduce the assessors' subjectivity.

Also, the evaluation process is carried out in a semi-quantitative way, where the attributes are evaluated qualitatively and the verbal expressions are, then, mapped into scores based on predefined guidelines [48]. Mapping these verbal descriptions into numeric numbers must be treated with more cautions.

Another issue that needs to be addressed in the future is that the reduced order-model is based on the fundamental assumption that the risk assessment model is correct (no model structural uncertainty). The reduced order model should be enhanced to consider the fact that the importance of the basic events depends on the structure of the risk assessment model itself. Finally, the output of the overall framework of risk aggregation is a risk distribution that accounts for the subjectivity in the analysis. The result cannot be used directly for comparison to the conventional single value-based safety criteria adopted in the current practice. Therefore, future work is needed for developing new safety criteria that correspond to risk estimates that consider trustworthiness, as well as developing guidelines for decision making support in the light of the outcomes of the developed framework.

References

[1] IAEA, *Determining the Quality of Probabilistic Safety Assessment (PSA) for Applications in Nuclear Power Plants*. Vienna: INTERNATIONAL ATOMIC ENERGY AGENCY, 2006.

[2] T. Aven and E. Zio, "Model output uncertainty in risk assessment," *Int. J. Performability Eng.*, vol. 9, no. 5, pp. 475–486, 2013.

[3] T. Bjerga, T. Aven, and E. Zio, "An illustration of the use of an approach for treating model uncertainties in risk assessment," *Reliab. Eng. Syst. Saf.*, vol. 125, pp. 46–53, 2014.

[4] NRC, *Reactor Coolant System and Connected Systems*. Washington: NRC, 2010.

[5] J. Eiser *et al.*, "Risk interpretation and action: A conceptual framework for responses to natural hazards," *Int. J. Disaster Risk Reduct.*, vol. 1, no. 1, pp. 5–16, 2012.

[6] R. Flage and T. Aven, "Expressing and communicating uncertainty in relation to quantitative risk analysis," *Reliab. Theory Appl.*, vol. 4, no. 2–1 (13), 2009.

[7] T. Bani-Mustafa, Z. Zeng, E. Zio, D. Vasseur, L. G. Industriel, and U. Paris-saclay, "Strength of Knowledge Assessment for Risk Informed Decision Making," in *Esrel*, 2018.

[8] T. Aven, "A conceptual framework for linking risk and the elements of the data-information-knowledge-wisdom (DIKW) hierarchy," *Reliab. Eng. Syst. Saf.*, vol. 111, pp. 30–36, 2013.

[9] Z. Zeng, F. Di Maio, E. Zio, and R. Kang, "A hierarchical decision-making framework for the assessment of the prediction capability of prognostic methods," *Proc. Inst. Mech. Eng. Part O J. Risk Reliab.*, vol. 231, no. 1, pp. 36–52, Dec. 2016.

[10] P. Kloprogge, J. P. Van der Sluijs, and A. C. Petersen, "A method for the analysis of assumptions in model-based environmental assessments," *Environ. Model. Softw.*, vol. 26, no. 3, pp. 289–301, 2011.

[11] C. Berner and R. Flage, "Strengthening quantitative risk assessments by systematic treatment of uncertain assumptions," *Reliab. Eng. Syst. Saf.*, vol. 151, pp. 46–59, 2016.

[12] Nasa, "STANDARD FOR MODELS AND SIMULATIONS-NASA-STD-7009," no. I, pp. 7–11, 2013.

[13] M. E. Paté-Cornell, "Uncertainties in risk analysis: Six levels of treatment," *Reliab. Eng. Syst. Saf.*, vol. 54, no. 2, pp. 95–111, 1996.

[14] G. Apostolakis, “The concept of probability in safety assessments of technological systems,” *Science (80-.).*, vol. 250, no. 4986, pp. 1359–1364, 1990.

[15] T. Askeland, R. Flage, and T. Aven, “Moving beyond probabilities ??? Strength of knowledge characterisations applied to security,” *Reliab. Eng. Syst. Saf.*, vol. 159, no. October 2016, pp. 196–205, 2017.

[16] T. Aven, “Improving risk characterisations in practical situations by highlighting knowledge aspects, with applications to risk matrices,” *Reliab. Eng. Syst. Saf.*, vol. 167, pp. 42–48, 2017.

[17] T. Aven, “Practical implications of the new risk perspectives,” *Reliab. Eng. Syst. Saf.*, vol. 115, pp. 136–145, 2013.

[18] T. Aven and B. S. Krohn, “A new perspective on how to understand, assess and manage risk and the unforeseen,” *Reliab. Eng. Syst. Saf.*, vol. 121, pp. 1–10, 2014.

[19] EPRI, “An Approach to Risk Aggregation for RisK-Informed Decision-Making,” Palo Alto, California, 2015.

[20] EPRI, “Guideline for the Treatment of Uncertainty in Risk-Informed Applications: Applications Guide,” Palo Alto, California: 1013491, 2006.

[21] J. P. Van Der Sluijs, M. Craye, S. Funtowicz, P. Kloprogge, J. Ravetz, and J. Risbey, “Combining Quantitative and Qualitative Measures of Uncertainty in Model-Based Environmental Assessment: The NUSAP System,” *Risk Anal.*, vol. 25, no. 2, pp. 481–492, 2005.

[22] I. Boone *et al.*, “NUSAP: a method to evaluate the quality of assumptions in quantitative microbial risk assessment,” *J. Risk Res.*, vol. 13, no. 3, pp. 337–352, 2010.

[23] A. De Jong, J. A. Wardekker, and J. P. Van der Sluijs, “Assumptions in quantitative analyses of health risks of overhead power lines,” *Environ. Sci. Policy*, vol. 16, pp. 114–121, 2012.

[24] E. Zio, “On the use of the analytic hierarchy process in the aggregation of expert judgments,” *Reliab. Eng. Syst. Saf.*, vol. 53, no. 2, pp. 127–138, 1996.

[25] C. Berner and R. Flage, “Quantitative vs. qualitative treatment of uncertain assumptions in risk assessment,” in *Safety and Reliability: Methodology and Applications - Proceedings of the European Safety and Reliability Conference, ESREL 2014*, 2015, pp. 2321–2328.

[26] J. Khorsandi and T. Aven, “Incorporating assumption deviation risk in quantitative risk assessments: A semi-quantitative approach,” *Reliab. Eng. Syst. Saf.*, vol. 163, pp. 22–32, 2017.

[27] E. L. Droguett and A. Mosleh, “Bayesian methodology for model uncertainty using model

performance data," *Risk Anal.*, vol. 28, no. 5, pp. 1457–1476, 2008.

[28] E. Lopez Droguett and A. Mosleh, "Bayesian Treatment of Model Uncertainty for Partially Applicable Models," *Risk Anal.*, vol. 34, no. 2, pp. 252–270, Feb. 2014.

[29] T. Bani-Mustafa, N. Pedroni, E. Zio, V. Dominique, and B. Francois, "A Hierarchical Tree-Based Decision Making Approach For Assessing The Trustworthiness Of Risk Assessment Models," in *2017 International Topical Meeting on Probabilistic Safety Assessment and Analysis (PSA 2017)*, 2017, pp. 314–323.

[30] T. Bani-Mustafa, Z. Zeng, E. Zio, and D. Vasseur, "A framework for multi-hazards risk aggregation considering risk model maturity levels," in *System Reliability and Safety (ICSRS), 2017 2nd International Conference on*, 2017, pp. 429–433.

[31] W. L. Oberkampf, M. Pilch, and T. G. Trucano, "Predictive capability maturity model for computational modeling and simulation," *cfwebprod.sandia.gov*, 2007.

[32] L. E. Schwer, "Guide for Verification and Validation in Computational Solid Mechanics," 2009.

[33] NRC, "AN APPROACH FOR USING PROBABILISTIC RISK ASSESSMENT IN RISK-INFORMED DECISIONS ON PLANT-SPECIFIC CHANGES TO THE LICENSING BASIS," 2011.

[34] Nicolas Zweibaum & Jean-Pierre Sursock, "Addressing multi-hazards risk aggregation for nuclear power plantsthrough response surface and risk visualization," Palo Alto, California, 2014.

[35] K. Simola and U. Pulkkinen, "Risk Informed Decision Making A Pre-Study," Nordisk Kernesikkerhedsforskning, Finland, 2004.

[36] EPRI, "Practical Guidance on the Use of Probabilistic Risk Assessment in Risk-Informed Applications with a Focus on the treatment of Uncertanity," Palo Alto, California, 2012.

[37] T. L. Saaty, *The Analytic Hierarchy Process*. 1980.

[38] M. C. Paulk, B. Curtis, M. B. Chrissis, and C. V Weber, "Capability Maturity Model for Software, Version 1.1," *Software, IEEE*, vol. 98, no. February, pp. 1–26, 1993.

[39] J. Herbsleb, D. Zubrow, D. Goldenson, W. Hayes, and M. Paulk, "Software quality and the capability maturity model," *Commun. ACM*, vol. 40, no. 6, pp. 30–40, 1997.

[40] F. Di Maio, P. Turati, and E. Zio, "Prediction capability assessment of data-driven prognostic methods for railway applications," in *Proceedings of the third European conference of the prognostic and health management society*, 2015.

[41] H. Veland and T. Aven, “Improving the risk assessments of critical operations to better reflect uncertainties and the unforeseen,” *Saf. Sci.*, vol. 79, pp. 206–212, 2015.

[42] T. Aven and B. Heide, “Reliability and validity of risk analysis,” *Reliab. Eng. Syst. Saf.*, vol. 94, no. 11, pp. 1862–1868, 2009.

[43] F. Goerlandt and J. Montewka, “Expressing and communicating uncertainty and bias in relation to Quantitative Risk Analysis,” *Saf. Reliab. Methodol. Appl.*, vol. 2, no. 13, pp. 1691–1699, 2014.

[44] L. Xu and J.-B. Yang, *Introduction to multi-criteria decision making and the evidential reasoning approach*. Manchester School of Management, 2001.

[45] E. Triantaphyllou and B. Shu, “Multi-criteria decision making: an operations research approach,” *Encycl. Electr. Electron. Eng.*, vol. 15, pp. 175–186, 1998.

[46] T. L. Saaty and L. G. Vargas, “How to make a decision,” in *Models, methods, concepts & applications of the analytic hierarchy process*, vol. 175, Springer Science & Business Media, 2012, pp. 1–20.

[47] M. Alexander, “Decision-Making using the Analytic Hierarchy Process (AHP) and SAS/ IML,” *United States Soc. Secur. Adm. Balt.*, pp. 1–12, 2012.

[48] T. L. Saaty, “Decision making with the analytic hierarchy process,” *Int. J. Serv. Sci.*, vol. 1, no. 1, p. 83, 2008.

[49] E. Mu and M. Pereyra-Rojas, “Understanding the analytic Hierarchy process,” in *Practical Decision Making*, Springer, 2017, pp. 7–22.

[50] E. Zio, M. Cantarella, and A. Cammi, “The analytic hierarchy process as a systematic approach to the identification of important parameters for the reliability assessment of passive systems,” *Nucl. Eng. Des.*, vol. 226, no. 3, pp. 311–336, 2003.

[51] T. L. Saaty and L. T. Tran, “On the invalidity of fuzzifying numerical judgments in the Analytic Hierarchy Process,” *Math. Comput. Model.*, vol. 46, no. 7–8, pp. 962–975, 2007.

[52] J. A. Alonso and M. T. Lamata, “Consistency in the analytic hierarchy process: a new approach,” *Int. J. Uncertainty, Fuzziness Knowledge-Based Syst.*, vol. 14, no. 4, pp. 445–459, 2006.

[53] I. Boone *et al.*, “A method to evaluate the quality of assumptions in quantitative microbial risk assessment,” *J. Risk Res.*, vol. 13, no. 3, pp. 337–352, 2010.

[54] R. Coudray and J. M. Mattei, “System reliability: An example of nuclear reactor system analysis,” *Reliab. Eng.*, vol. 7, no. 2, pp. 89–121, 1984.

[55] P. H. A. M. T. Unwin SD, PP Lowry, RF Layton, Jr, “Multi-State Physics Models of Aging Passive Components in Probabilistic Risk Assessment,” in *In International Topical Meeting on Probabilistic Safety Assessment and Analysis*, 2011, p. vol. 1, pp. 161–172.

[56] F. Di Maio, D. Colli, E. Zio, L. Tao, and J. Tong, “A multi-state physics modeling approach for the reliability assessment of nuclear power plants piping systems,” *Ann. Nucl. Energy*, vol. 80, pp. 151–165, 2015.

[57] R. D. Burns, “Wash 1400—Reactor safety study,” *Prog. Nucl. Energy*, vol. 6, no. 1–3, pp. 117119–117140, 1980.

[58] T. Aven, “On the use of conservatism in risk assessments,” *Reliab. Eng. Syst. Saf.*, vol. 146, pp. 33–38, 2016.

[59] H. Riesch, “Uncertainty,” in *Essentials of Risk Theory*, no. 2009, 2013, pp. 29–57.

[60] R. Ferdous, F. Khan, R. Sadiq, P. Amyotte, and B. Veitch, “Analyzing system safety and risks under uncertainty using a bow-tie diagram: An innovative approach,” *Process Saf. Environ. Prot.*, vol. 91, no. 1, pp. 1–18, 2013.

[61] H. Abdo, J. M. Flaus, and F. Masse, “Uncertainty quantification in risk assessment-Representation, propagation and treatment approaches: Application to atmospheric dispersion modeling,” *J. Loss Prev. Process Ind.*, vol. 49, pp. 551–571, 2017.

[62] W. E. Walker *et al.*, “Defining Uncertainty: A Conceptual Basis for Uncertainty Management in Model-Based Decision Support,” *Integr. Assess.*, vol. 4, no. 1, pp. 5–17, 2003.

[63] B. Wynne, “Uncertainty and environmental learning: Reconceiving science and policy in the preventive paradigm,” *Glob. Environ. Chang.*, vol. 2, no. 2, pp. 111–127, 1992.

[64] D. J. Spiegelhalter and H. Riesch, “Don’t know, can’t know: embracing deeper uncertainties when analysing risks,” *Phil. Trans. R. Soc. A*, vol. 369, no. 1956, pp. 4730–4750, 2011.

[65] W. K. Viscusi, J. T. Hamilton, and P. C. Dockins, “Conservative versus mean risk assessments: Implications for Superfund policies,” *J. Environ. Econ. Manage.*, vol. 34, no. 3, pp. 187–206, 1997.

[66] R. M. Perhac Jr, “Does risk aversion make a case for conservatism,” *Risk*, vol. 7, p. 297, 1996.

[67] C. Whipple, “Dealing with uncertainty about risk in risk management,” in *Risk Assessment and Management*, Springer, 1987, pp. 529–536.

[68] M. Davies, “Knowledge–Explicit, implicit and tacit: Philosophical aspects,” *Int. Encycl. Soc. Behav. Sci.*, pp. 74–90, 2015.

[69] T. Bjerga and T. Aven, “Adaptive risk management using new risk perspectives – an example from the oil and gas industry,” *Reliab. Eng. Syst. Saf.*, vol. 134, pp. 75–82, 2015.

[70] T. Aven and M. Ylönen, “Safety regulations: Implications of the new risk perspectives,” *Reliab. Eng. Syst. Saf.*, vol. 149, pp. 164–171, 2016.

[71] S. D. Talisayon, “Monitoring and evaluation in knowledge management for development,” *IKM Emergent Pap.*, vol. 3, 2009.

[72] D. G. Cacuci, M. Ionescu-Bujor, and I. M. Navon, *Sensitivity and uncertainty analysis*, vol. 1. Chapman & hall/CRC Boca Raton, Florida, 2003.

[73] H. Christopher Frey and S. R. Patil, “Identification and review of sensitivity analysis methods,” *Risk Anal.*, vol. 22, no. 3, pp. 553–578, 2002.

[74] E. Borgonovo and A. Cillo, “Deciding with Thresholds: Importance Measures and Value of Information,” *Risk Anal.*, vol. 37, no. 10, pp. 1828–1848, 2017.

[75] E. Zio and N. Pedroni, *Overview of risk-informed decision-making processes*. FonCSI, 2012.

[76] J. M. Reinert and G. E. Apostolakis, “Including model uncertainty in risk-informed decision making,” *Ann. Nucl. Energy*, vol. 33, no. 4, pp. 354–369, 2006.

[77] D. M. Hamby, “A review of techniques for parameter sensitivity analysis of environmental models,” *Environ. Monit. Assess.*, vol. 32, no. 2, pp. 135–154, 1994.

[78] D. J. Downing, R. H. Gardner, and F. O. Hoffman, “An examination of response-surface methodologies for uncertainty analysis in assessment models,” *Technometrics*, vol. 27, no. 2, pp. 151–163, 1985.

[79] E. Zio, “Basic of Probability Theory for Applications to Reliability and Risk Analysis,” in *An introduction to the basics of reliability and risk analysis*, vol. 13, World scientific, 2007, pp. 31–32.

[80] RELCON AB, *RiskSpectrum Professional: Theory Manual*. 2005.

[81] R. Koch, *The 80/20 principle: the secret to achieving more with less*. Hachette UK.: Crown Business, 2011.

[82] IAEA, “Development and Application of Level 1 Probabilistic Safety Assessment for Nuclear Power Plants,” 2010.

[83] IAEA, “External Events Excluding Earthquakes in the Design of Nuclear Power Plants,” 2003.

[84] IAEA, “Deterministic Safety Analysis for Nuclear Power Plants,” 2009.

[85] M. Knochenhauer and J. E. Holmberg, “Guidance for the definition and application of probabilistic

safety criteria," *Proc. PSAM 10 Int. Probabilistic Saf. Assess. Manag.*, 2012.

[86] IAEA, "Meteorological and Hydrological Hazards in Site Evaluation for Nuclear Installations," 2009.

[87] E. L. Droguett, "Methodology for the treatment of model uncertainty," University of Maryland at College Park., 1999.

[88] 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*, 2010, pp. 8-p.

[89] L. Jiao, Q. Pan, Y. Liang, X. Feng, and F. Yang, "Combining sources of evidence with reliability and importance for decision making," *Cent. Eur. J. Oper. Res.*, vol. 24, no. 1, pp. 87–106, 2016.

[90] J. Dezert and J.-M. Tacnet, "Evidential reasoning for multi-criteria analysis based on DSmT-AHP," *Adv. Appl. DSmT Inf. Fusion*, p. 95, 2011.

[91] A. H. Tayyebi, M. R. Delavar, A. Tayyebi, and M. Golobi, "Combining multi criteria decision making and Dempster Shafer theory for landfill site selection," *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci*, vol. 38, no. 8, pp. 1073–1078, 2010.

[92] A. Ennaceur, Z. Elouedi, and E. Lefevre, "Handling partial preferences in the belief AHP method: Application to life cycle assessment," in *Congress of the Italian Association for Artificial Intelligence*, 2011, pp. 395–400.

[93] M. Beynon, D. Cosker, and D. Marshall, "An expert system for multi-criteria decision making using Dempster Shafer theory," *Expert Syst. Appl.*, vol. 20, no. 4, pp. 357–367, 2001.

[94] T. L. Saaty, "Analytic hierarchy process," in *Encyclopedia of operations research and management science*, Springer, 2013, pp. 52–64.

[95] G. Shafer, *A mathematical theory of evidence*, vol. 42. Princeton university press, 1976.

[96] P. Smets and R. Kennes, "The transferable belief model," *Artif. Intell.*, vol. 66, no. 2, pp. 191–234, 1994.

[97] A. Aregui and T. Denœux, "Constructing consonant belief functions from sample data using confidence sets of pignistic probabilities," *Int. J. Approx. Reason.*, vol. 49, no. 3, pp. 575–594, 2008.

[98] F. Groen and A. Mosleh, "Behavior of weighted likelihood and weighted posterior methods for treatment of uncertain data," in *Proc. ESREL*, 1999, vol. 99.

[99] S. Kaplan and B. J. Garrick, “On the quantitative definition of risk,” *Risk Anal.*, vol. 1, no. 1, pp. 11–27, 1981.

[100] R. Kazemi and A. Mosleh, “Improving default risk prediction using Bayesian model uncertainty techniques,” *Risk Anal. An Int. J.*, vol. 32, no. 11, pp. 1888–1900, 2012.

[101] D. Jenkinson, “The elicitation of probabilities: A review of the statistical literature,” Citeseer, 2005.

[102] IAEA, “IAEA-Publication8635.” 2011.

[103] U. S. NRC, “General Electric Advanced Technology Manual Chapter 4.8 Service Water System Problems,” 2011.

[104] T. Gschnitzer, B. Gems, B. Mazzorana, and M. Aufleger, “Towards a robust assessment of bridge clogging processes in flood risk management,” *Geomorphology*, vol. 279, pp. 128–140, 2017.

[105] R. D. J. M. Steenbergen, P. van Gelder, S. Miraglia, and A. Vrouwenvelder, “Safety, reliability and risk analysis: beyond the horizon,” 2013, pp. 3355–3361.

Part (II)
Appendixes

Appendix I (P1): A hierarchical tree-based decision making approach for assessing the relative trustworthiness of risk assessment models

A hierarchical tree-based decision making approach for assessing the relative trustworthiness of risk assessment models

Tasneem Bani-Mustafa1[(1)], Nicola Pedroni[(2)], and Enrico Zio[(1), (3)], Dominique Vasseur[(4)] & Francois Beaudouin[(5)]

[(1)] *Chair on System Science and the Energetic Challenge, EDF Foundation*

Laboratoire Genie Industriel, CentraleSupélec/Université Paris-Saclay, 3 Rue Joliot Curie, 91190 Gif-sur-Yvette, France , tasneem-adeeb.bani-mustafa@centralesupelec.fr

[(2)] *Politecnico di Torino, Corso Duca degli Abruzzi, 24 - 10129 Torino (Italy)*

[(3)] *Energy Department, Politecnico di Milano, Via Ponzio 34/3, Milan, 20133, Italy*

[(4)] *EDF R&D, PERICLES (Performance et prévention des Risques Industriels du parc par la simulation et les Etudes)*

EDF Lab Paris Saclay - 7 Bd Gaspard Monge, 91120 Palaiseau, France

[(5)] *EDF R&D, PRISME (Performance, Risque Industriel, Surveillance pour la Maintenance et l'Exploitation)*

EDF Lab Chatou - 6 Quai Watier 78401 Chatou

Abstract:

Risk assessment provides information to support Decision Making (DM). Then, the confidence that can be put in its outcomes is fundamental, and this depends on the accuracy, representativeness and completeness of the models used in the risk assessment. Some sort of quantitative measure must be provided to assess the credibility and trustworthiness of the results obtained from such models for DM purposes.

The present paper proposes a four-levels, top-down, hierarchical tree to identify the main attributes and criteria that affect the level of trustworthiness of models used in probabilistic risk assessment. The level of trustworthiness is broken down into two attributes (Level 2), three sub-attributes (Level 3), one "leaf" attribute (Level 3), and seven basic "leaf" sub-attributes (Level 4). On the basis of this hierarchical decomposition, a bottom up, quantitative approach is employed for the assessment of model trustworthiness, using tangible information and data available for the basic "leaf" sub-attributes (Level 4). The analytical hierarchical process (AHP) is adopted for evaluating and aggregating the sub-attributes.

The approach is applied to a case study concerning the modeling of the Residual Heat Removal (RHR) system of a nuclear power plant (NPP), to compute its failure probability. The relative trustworthiness of two mathematical models of different complexity is evaluated: a Fault Tree (FT) and a Multi-States Physics-based Model (MSPM). The feasibility and reasonableness of the approach are demonstrated, paving the way for its potential applicability to inform DM on safety-critical systems.

Keywords:

Risk assessment, Risk-Informed Decision Making (RIDM), Strength of Knowledge, Model Trustworthiness and Credibility, Fault tree, Multi-States Physics-Based Model (MSPM), Analytical Hierarchical Process (AHP), Residual Heat Removal (RHR) System, Nuclear Power Plant (NPP).

1. Introduction

Risk assessment is based on models that represent the functional life and physical behavior of (safety-critical) systems and processes of interest. These models are conceptual constructs (translated into mathematical forms), built on a set of assumptions (hypotheses) made on the basis of the available knowledge. In this sense, the risk assessment outcomes are conditional on the available knowledge. Then, the qualitative risk insights and quantitative risk indices drawn from the risk assessment may have a more or less solid foundation, depending on the validity of the hypotheses made, which in turn depends on the supporting knowledge.

In general terms, risk describes the future *consequences* (usually seen in negative, undesirable terms with respect to the planned objectives) potentially arising from the operation of given systems and activities, and the associated *uncertainty* (INSAG 2011). Risk should, then, be qualitatively described and quantitatively assessed in order to compare it with predefined safety criteria, for further guidance to risk-informed decision making (RIDM) (Dezfuli et al. 2010); (NRC 2010); (Eiser et al. 2012).

Risk assessments rely on the use of complex *models* to represent systems and processes, and provide predictions of safety performance metrics (Aven & Zio 2013). These models are (interpreted and simplified) conceptual constructs (translated into mathematical forms) built on a set of assumptions (hypotheses).

In recent times, there have been a vivid discussion on the fundamental concept of "risk" and related foundational issues on its assessment: (see, e.g., (Aven 2013a); (Aven 2016); (Cox & Lowrie 2015)). From a general perspective, it is understood that the outcomes of risk assessments (i.e., the undesirable events/scenarios, consequences and the description of uncertainty about these) are conditioned on the *background knowledge* and *information* available on the system and/or process under analysis (Bjerga et al. 2014); (Zeng et al. 2016), including assumptions and presuppositions, phenomenological understanding, historical system performance data and expert statements used (Flage & Aven 2009); (Aven 2013b) (Veland & Aven 2015); (Berner & Flage 2016); (Bani-mustafa et al. 2018). Then, the risk indices may have a more or less solid foundation, depending on the validity of the hypotheses made: poor models, lack of data or simplistic assumptions are examples of potential sources of (model) uncertainty "hidden in the background knowledge" of a risk assessment (Berner & Flage 2016). The modeling of a system or process needs to balance between two conflicting concerns: (i) *accurate representation* of the phenomena and mechanisms in the system or process and (ii) definition of the proper *level of detail* of the description of the phenomena and mechanisms, so as to allow the timely and efficient use of the model. Differences between the real world quantities and the model outputs inevitably arise from the conflict of these two concerns (Paté-Cornell 1996); (Bjerga et al. 2014); (Danielsson *et al.*, 2016). Since (i) the importance placed on modeling and simulation is increasingly high within safety-critical system engineering contexts and (ii) the fundamental value of a risk assessment lies in providing informative support to (high-consequence) decision making (DM) (Simola & Pulkkinen 2004); (EPRI 2012); (Eiser et al. 2012); (Zweibaum & Sursock, 2014), the *confidence* that can be put in the accuracy, representativeness and completeness of the models is fundamental and a satisfactory level of assurance must be provided that the

results obtained from such models are *credible* and *trustworthy* for the decision-making purposes for which they are employed. Moreover, in some contexts where the system of interest is subject to multiple hazards (e.g., a Nuclear Power Plant (NPP) exposed to floodings and earthquakes risks), a Multi-Hazards Risk Aggregation (MHRA) process is required to obtain a final risk metric that can inform decision making. However, risk estimates for different (risk) contributors are typically obtained using different models (i.e., in practice, different PRAs), each one having its own level of maturity and relying on its background knowledge. This inconsistency might be problematic, as MHRA is often carried out by a simple arithmetic summation of the risk estimates from different contributors, ignoring the possibly different levels of knowledge, which the risk estimates are based on (EPRI 2015). Another situation, where the use of risk models with different credibility might be problematic, is that of choosing between the implementation of two different sets of risk reduction measures. For example, in a purely RIDM, a decision maker would always choose the option leading to the lower level of risk; however, his/her decision could change if he/she considered the level of trustworthiness, which the corresponding risk estimates are based on. For all these reasons, the *confidence*, *credibility* and *trustworthiness* (resp., *model uncertainty*) that is associated with model predictions (and that reflects the *amount* and the *strength* of the *knowledge* available on the problem of interest), must be accurately and quantitatively assessed (Aven & Zio 2013); (Bjerga et al. 2014); (Flage & Aven 2015).

Within this context, the objective of the present paper is to propose a four-levels, top-down, hierarchical tree-based decision-making approach to assess the relative trustworthiness of different models used in a given risk assessment. On the other hand, it is out of the scope of the present paper to propose a general framework to integrate the level of trustworthiness in Risk Informed Decision Making (RIDM) process. In this framework, the level of trustworthiness is divided into two attributes (level 2), four sub-attributes (level 3) and seven basic "leaf" sub-attributes (level 4). The alternative models whose trustworthiness and credibility is to be assessed all at the bottom of the structure. On the basis of this hierarchical decomposition, the level of trustworthiness is, then, calculated by resorting to a bottom-up, quantitative approach. The basic "leaf" attributes represent tangible attributes that can be directly and quantitatively evaluated using data and information available (e.g., past knowledge, experts judgments, historical records, etc.). In the present study, the Analytical Hierarchical Process (AHP) is employed for evaluating and aggregating (in weighted fashion) the sub-attributes.

The proposed approach has been applied to assess the trustworthiness of two models (of different complexity and level of detail) of a Residual Heat Removal (RHR) System of the ~~–~~Nuclear Power Plant (NPP)~~:~~ the two models are used to estimate the failure probability of the safety system of interest. The first model is based on a classical Boolean logic-based Fault Tree (FT). This approach employs components' failure rates that are simply based on field data and/or expert judgment. The model does not consider possible dependencies existing between the states of degradation of different components (e.g., a valve and a pump) nor the interaction between physical and environmental parameters and the mechanisms of components' degradation (Lin, 2016). On the other hand, the second approach is based on a Multi-States Physics-based Model (MSPM), which takes into account multiple time-dependent components'

degradation states, the effects of physical and environmental parameters on the mechanisms of degradation, and the dependencies between the degradations of components (Unwin *et al.*, 2011); (Lin *et al.*, 2013); (Lin et al. 2015); (Lin *et al.*, 2016).

A review of the approaches proposed in the literature to assess the trustworthiness and credibility of a model is presented in Section 2. In Section 3, a hierarchical tree-based decision making framework for assessing model trustworthiness is presented. In Section 4, the proposed framework is applied to a case study concerning the RHR system of a NPP. Finally, in Section 5, we discuss the results and provide some conclusions.

2. Assessing the trustworthiness and credibility of risk assessment models: a critical review of literature

In this section, we survey some approaches proposed in the open literature to assess the trustworthiness and credibility of mathematical models.

Few methods have been proposed to assess the confidence (i.e., the credibility and trustworthiness) that is associated with engineering model predictions and that reflects the amount and the strength of the knowledge available on a generic system, or process of interest. In the literature, the trustworthiness of a method or a process is often measured in terms of its maturity. The concept of a model maturity goes back to the 1970s: at the time, it was used to assess the maturity of a function of an information system (Oberkampf *et al.*, 2007); (Paulk *et al.*, 1993); (Zeng et al. 2016). Later, the Software Engineering Institute (SEI) developed a framework (the so-called Capability Maturity Model (CMM)) to assess the maturity of a software development process, in the light of its quality, reliability and trustworthiness (Herbsleb *et al.*, 1997). Recently, the CMM model has been extended and a Prediction Capability Maturity Model (PCMM) has been developed to evaluate and assess the maturity of modeling and simulation efforts (Oberkampf et al. 2007). Other examples of maturity assessment approaches have been developed in different domains, such as master data maturity assessment, enterprise risk management and hospital information system (Zeng et al. 2016). In (Di Maio *et al.*, 2015) and (Zeng et al. 2016) a hierarchical framework based on the analytical hierarchical process (AHP) has been developed to assess the maturity and prediction capability of a prognostic method for maintenance DM purposes. Finally, a framework for assessing the credibility of models and simulation (M&S) is proposed by (Nasa 2013). In this framework, eight factors are used to assess the credibility of Models & Simulation (M&S) and are categorized in three groups: (i) M&S development including verification and validation; (ii) M&S operations, including input pedigree, results uncertainty and results robustness; (iii) supporting evidence, including the use history, M&S management and people qualifications. This framework seems plausible and covers important elements. However, three main issues should be considered: first, the approach is abstractly presented, leading to omit some important elements that fall under the main attributes of this framework. For example, while the model focuses on the "input pedigree" represented by the input data, it ignores a very important element, i.e., model assumptions, that can be also a part of M&S development. Second, while the authors claim that there is no need for weighting the elements, as there is no numerical aggregation required, this would lead to a misconception, since the elements are not equally important in practice. For example, at a first glance,

one would consider “use history” as important as “validation”, but the attribute “validation”, which checks the accuracy of the model’s representation of the real system, may be considered more important than “use history” as using a model for a long time does not necessarily guarantee that it gives good results and, thus, better informed decisions (e.g., a model could be adopted because of its simplicity or because its use is motivated by an “established tradition” within a given community).

In the more specific field of “strength of knowledge” assessment in risk assessment models, both qualitative and semi-quantitative approaches have been proposed. In (Flage & Aven 2009), a “crude” qualitative, direct grading of the strength of knowledge that supports risk assessment based on (mathematical) models is introduced. The authors try to classify the strength of knowledge to {minor, moderate, significant}, with respect to the following elements (Flage & Aven 2009); (Berner & Flage 2016); (Aven 2013b); (Veland & Aven 2015); (Bani-mustafa et al. 2018):

1. phenomenological understanding of the problem and availability of precise and well-understood predicting models for the physical phenomena of interest;
2. availability of reliable data;
3. reasonability of assumptions made (i.e., the assumptions do not exhibit large simplifications);
4. agreement (consensus) among experts (i.e., low value ladenness).

The strength of knowledge is, then, classified according to the following criteria (Flage & Aven 2009); (Berner & Flage 2016); (Aven 2013b); (Veland & Aven 2015); (Bani-mustafa et al. 2018):

1. if none of the previously mentioned components is met, then the knowledge is “weak”;
2. if the “requirements” are partially met, then the strength of knowledge is considered “intermediate”;
3. if all “requirements” are met, then, the knowledge is considered “strong”.

In (Aven 2013b) a more detailed, semi-quantitative approach (namely the “assumption deviation risk”) has been introduced. This approach is based on the identification of all the main assumptions on which the analysis is based. Then, the assumptions are converted into uncertainty factors and a rough evaluation of the deviation from the conditions defined by the assumptions is carried out. Finally, a score is assigned to each deviation that reflects the risk related to the deviation and its implications on the occurrence of given events and their consequences. Notice that the score captures all the components of the risk concept, i.e., the deviation from the assumptions made with the associated consequences, the uncertainty of this deviation and consequences, and the strength of knowledge that these are based on (Aven 2013b); (Berner & Flage 2016).

In (Berner & Flage 2016), the authors embrace, apply, test and adjust the perspectives of (Flage & Aven 2009) and (Aven 2013b) to develop a general and systematic framework for treating (uncertain) assumptions in risk assessment models. Also, this methodology for assessing the importance of assumptions is based on evaluating the basic elements of the risk description mentioned above and previously developed and adopted by (Aven 2013b). The evaluation places an assumption in one of six “settings”, each providing guidelines for characterizing the corresponding uncertainty. In practice, these guidelines and strategies are based on the precept that the effort that should be exerted for characterizing

the uncertainty associated to an assumption and the effect on it of the potential deviations, should increase with the importance and the criticality of the assumption.

Also in (Bjerga et al. 2014) the effect and importance of "structural" assumptions, approximations and simplifications on risk assessment model outputs (Aven & Zio 2013) is studied by means of different approaches, including subjective and imprecise probabilities and semi-quantitative scores (reflecting the degree of uncertainty associated to an assumption and the sensitivity of the model output to such assumption). The analysis serves as an input to the decision makers, to understand which assumptions are unacceptable and need "remodeling".

Finally, Lopez-Droguett and Mosleh discuss uncertainty in model predictions arising from model parameters and the model structure. They argue that different evidence in evaluating model uncertainty can be considered, such as: comparing the results of the model predication to the actual measurements, qualitative or subjective evaluation of the model credibility and applicability (Droguett & Mosleh 2008). In particular, for cases in which no model exists to address the particular problem of interest, and the analysis rely mainly on the subjective assumptions that the model is partially applicable to the problem, two main attributes define model uncertainty: model *Credibility* and model *Applicability* (Lopez Droguett & Mosleh, 2014). Model credibility refers to the quality of the model in estimating the unknown in its intended domain of application and is defined by a set of attributes related to the model-building process and utilization procedure (*conceptualization and implementation, which are in turn broken down into other sub-attributes). On* the other hand, model applicability represents the degree to which the model is suitable for the specific situation and problem (represented by the *conceptualization* and *intended use function* attributes) (Lopez Droguett & Mosleh, 2014).

3. Hierarchical tree-based decision making approach for assessing the trustworthiness of risk assessment models

In section 3.1 below, we present the four levels, top-down tree used to characterize the trustworthiness (of a risk assessment model) by decomposing it into sub-attributes (e.g., number of model's assumptions, quantity of relevant data available, etc.) that can be quantified by the analysts; in Section 3.2, we describe a bottom-up procedure, based on the analytical hierarchal process (AHP), to assess the model trustworthiness by evaluating and aggregating the sub-attributes (identified as "leaf" attributes).

3.1. Hierarchical tree for model trustworthiness characterization: abstraction and decomposition

Many factors (attributes) affect the trustworthiness and credibility of analyses and models (for risk assessment in particular), and several studies and literature reviews have been made in order to identify them. Some of these are summarized as follows: (i) phenomenological understanding of the problem; (ii) availability of reliable data; (iii) reasonability of the assumptions; (iv) agreement among the experts; (v) level of detail in the description of the phenomena and processes of interest; (vi) accuracy and precision in the estimation of the values of the model's parameters; (vii) level of conservatism; (viii) amount of uncertainty and others (see e.g., (Flage & Aven 2009); (Berner & Flage 2016); (Aven 2013a); (Veland & Aven 2015); (IAEA, 2006); (Bjerga et al. 2014); (Zeng et al. 2016); (Oberkampf et al. 2007); (EPRI 2012);

(EPRI 2015); (Bani-mustafa et al. 2018)). Some of these attributes (criteria), are not tangible and cannot be measured directly: as a consequence, other sub-attributes must be identified, which can be measured and/or subjectively evaluated. To this aim, on the basis of the critical literature survey presented in Section 2, we propose a method for model trustworthiness characterization and decomposition, which is based on the hierarchy tree shown in Figure 1.

As mentioned above, many factors can be found in the literature that characterize the level of trustworthiness. Those factors can be categorized into two main groups: (i) "strength of knowledge"; (ii) "modeling fidelity", which embody the ability of a model of representing the reality and the degree of implementing correctly the model. In the "strength of knowledge", among the four sub-elements proposed in (Flage & Aven 2009), two were found to be more relevant to the context of interest i.e., data and assumptions. In the latter, it is argued that including more details about a problem is more representative and realistic, and hence more trustworthy. For example, there are different levels of PRA treatment that are chosen, relying on alternative decisions (Paté-Cornell 1996). On the other hand, implementing the model correctly from a pure trustworthiness point of view, without considering a costs-benefits reasoning, requires avoiding approximation. In accordance, a hierarchical tree for models' trustworthiness is proposed in Figure 1.

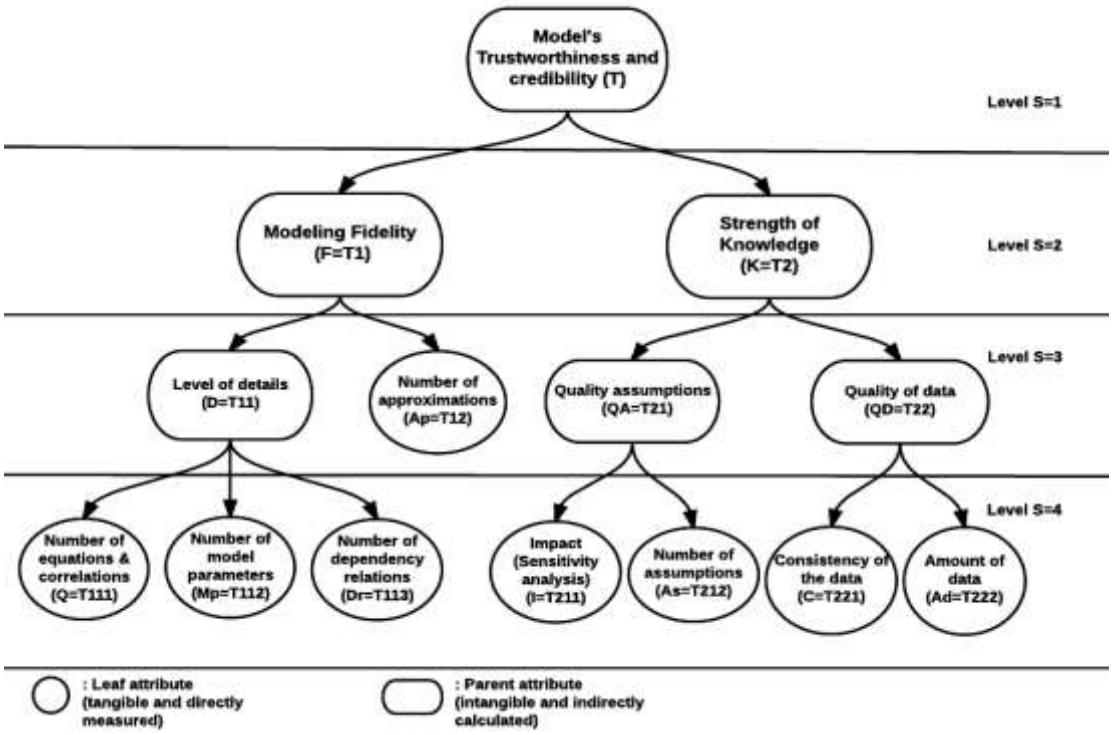


Figure 1 A hierarchical tree-based "decomposition" of the level of trustworthiness and credibility of a mathematical model

The model trustworthiness, represented by T (Level 1), is characterized by two attributes: modeling fidelity, represented by $F = T_1$ and strength of knowledge, represented by $K = T_2$ (Level 2). The modeling fidelity ($F = T_1$), measures the adequacy of the model representation of the phenomenon and the

level of detail adopted in the model description (referred to as modeling validity in some literatures (Aven & Heide 2009)).On the other hand, the strength of knowledge ($K = T_2$) measures how solid the assumptions, data and information (which the model relies on) are (Flage & Aven 2009). These two attributes are in turn decomposed into sub-attributes (Level 3). In particular, the modeling fidelity $F = T_1$ is defined by the level of detail, represented by $D = T_{11}$ (Level 3) and by the number of approximations, represented by $Ap = T_{12}$. Concerning the strength of knowledge $K = T_2$, among the four sub-attributes proposed in Flage & Aven (2009) (see Section 2), two are found to be more relevant to the context indeed, i.e. data and assumptions. Thus, attribute $K = T_2$ is here defined by the quality of assumptions represented by $QA = T_{21}$ and by the quality of data represented by $QD = T_{22}$. Note that the number of approximations $Ap = T_{12}$ is considered as a basic attribute, since it can be measured directly: thus, it is not further broken down into other sub-attributes. The other three attributes of Level 3 are instead broken down into more basic "leaf" attributes that can be measured directly by "inspection" of the model whose trustworthiness we want to assess. In particular, the level of detail $\mathrm{D} = \mathrm{T}_{11}$ is characterized in terms of the number of equations and correlations, namely $\mathrm{Q} = \mathrm{T}_{111}$, the number of model parameters, namely $\mathrm{Mp} = \mathrm{T}_{112}$, and the number of dependency relations included, namely $\mathrm{Dr} = \mathrm{T}_{113}$. The overall quality of the assumptions $\mathrm{QA} = \mathrm{T}_{21}$ is measured by the number of assumptions made $\mathrm{As} = \mathrm{T}_{212}$, and by their impact $\mathrm{I} = \mathrm{T}_{212}$ (which can be assessed, e.g., by sensitivity analysis). Finally, the quality of the data $\mathrm{QD} = \mathrm{T}_{22}$ is described in terms of the amount of data available, namely $\mathrm{Ad} = \mathrm{T}_{221}$ and by the consistency of the data itself, namely $\mathrm{C} = \mathrm{T}_{222}$. Precise definitions of the attributes are given in Table 1 for the sake of clarity.

Table 1 Definition of the attributes used to characterize the model trustworthimness

Level	Attributes	Description
Level S = 2	Modeling fidelity $F = T_1$	Measures how close the model is to reality, i.e., the adequacy of the representation of the phenomena and processes of interest: the higher the modeling fidelity, the higher the trustworthiness of the model.
	Strength of knowledge $K = T_2$	Represents the level of understanding of the phenomena and the solidity of the assumptions, data and information, which the model relies on: the higher the strength of knowledge, the higher the trustworthiness of the model.
Level S = 3	Level of detail $D = T_{11}$	Measures the level of sophistication of the analysis by quantifying to which level the "elements" and aspects of the phenomenon, process or system of interest are taken into account in the model: the higher the level of detail, the higher the trustworthiness of the model.
	Number of approximations $Ap = T_{12}$	Measures the number of approximations that the analyst introduces in order to facilitate the analysis: it affects the modeling fidelity. The lower the number of model approximations the higher the modeling fidelity.
	Quality of assumptions $QA = T_{21}$	In some studies, experts are obliged to formulate some assumptions, which might be due to the lack of data and information, to the complexity of the problem or to lack of phenomenological understanding. The quality of those assumptions is an indication of the strength of knowledge: the higher the quality of the assumptions, the higher the trustworthiness of the model.

	Quality of data $QD = T_{22}$	Represents the availability of sufficient, accurate and consistent background data with respect to the purposes of the analysis: the higher the quality of the data, the higher the trustworthiness of the model.
Level S = 4	Number of equations and correlations $Q = T_{111}$	The number of equations and correlations used in modeling is an indication of the level of detail, hence of the modeling fidelity: the higher the number of equations and correlations, the higher the trustworthiness of the model.
	Number of model parameters $Mp = T_{112}$	The number of parameters introduced in the model is a measure of the level of detail (e.g., the number of components transition rates represents the level of discretization adopted to describe the failure process of a component or a system): the higher the number of model parameters, the higher the trustworthiness of the model.
	Number of dependency relations $Dr = T_{113}$	The larger the number of dependency relations that are taken into account, the more detailed and trustworthy the model.
	Number of assumptions $As = T_{211}$	The larger the number of high quality assumptions, the higher the trustworthiness of the model.
	Impact of assumptions $I = T_{212}$	It quantifies how much assumptions can affect the model results (and it can be assessed by sensitivity analysis). The higher the impact of the assumptions, the lower the trustworthiness of the model.
	Consistency of data $C = T_{221}$	It is an indication of how suitable and representative the data are for a specific process or system. The consistency of data relies on the sources of the data. For example, if we are collecting data about the failure of a safety system's pump from different power plants, we should first understand whether the power plants are of the same type, whether the plants work at the same power level and whether the pumps have the same work function and capacity. The consistency of the data used is an indication of the quality of data, hence of the strength of knowledge: the higher the consistency, the higher the strength of knowledge and the trustworthiness of the model.
	Amount of data $Ad = T_{222}$	The higher the amount of data available, the stronger the knowledge. For example, the number of years of experience of a particular component in a plant can be sometimes considered an indication of the amount of data available. In any domain, a higher number of years' experience means a higher number of scenarios covered and hence a larger amount of data. The higher the amount of data, the higher the trustworthiness of the model.

Some considerations are in order with respect to the hierarchical decomposition described above. There is no claim that the approach proposed is comprehensive and complete, since other attributes may affect model credibility and, hence, trustworthiness. For example, an increase in the number of parameters of a model, on one side, increases the level of details that the model is capable to capture but, on the other side, it may leave room for additional errors and uncertainties in its estimated parameters (which are not included in the present formulation). As specified before, the constituting attributes have been selected on the basis of an accurate and critical literature review of works treating the subject (see Section 2). Also, guidelines have been developed to provide a quantitative (or semi-quantitative) evaluation of such elements. These guidelines have been developed on the basis of the experience and knowledge of EDF

experts (see Appendix A). So, the contribution has to be considered a first attempt of a systematic framework to address the evaluation of model trustworthiness and to give a structure to expert judgment on this, which is absolutely inevitable in this type of analysis.

3.2. Analytical hierarchical process (AHP) for model trustworthiness quantification

Given the hierarchical tree in Figure 1, the assessment of model trustworthiness is carried out within a multi-criteria decision analysis (MCDA) framework (Xu & Yang 2001); (Triantaphyllou & Shu 1998). In this setting, we suppose, in all generality, that a system, process or phenomenon of interest for a risk assessment can be represented by different mathematical models of possibly different complexity and level of detail, $M1, M2, .. M_l, .., M_n$. The task (i.e., the MCDA problem at hand) is to rank these alternative models with respect to their trustworthiness, in relation to the particular risk assessment problem of interest to support MCDA. In the present paper, the Analytical Hierarchy Process (AHP) proposed by (Saaty & Vargas 2012) is adopted to this aim. However, other MCDA approaches could be used.

In this approach, the top goal, i.e., the decision problem considered (in this case, the model trustworthiness), is placed at the first level of the hierarchy and, then, decomposed into several sub-attributes distributed over different levels according to their degree of tangibility. Finally, the bottom level in the hierarchal tree-based AHP model contains the different alternatives that need to be evaluated with respect to the top goal (i.e., in this case the level of trustworthiness) (Saaty 2008); (Zeng et al. 2016). Through pairwise comparisons among the elements and the attributes of the same level, the alternative solutions, i.e., models, can be ranked with respect to the decision problem in the top level (i.e., the model trustworthiness) (Saaty 2008); (Zio *et al.*, 2003). A good feature of the method is that it can be helpful in group-decision-making (Saaty 2008), and in situations that involve mixed quantitative and qualitative factors (Alexander 2012).

The AHP model for model trustworthiness assessment is represented in Figure 1. The first step required to assess the model trustworthiness by AHP is the determination of the so-called inter-level priorities (in practice, weights that represent the importance of attributes in the same level relative to their parent attribute) for each attribute, sub-attribute, basic "leaf" sub-attribute and alternative solution i.e., $W(T_i)$, $W(T_{ij})$, $W(T_{ijk})$, and $W(M_l, T_{ijk})$, respectively. Notice that in practice, each weight represents the relative contribution of an attribute of a given level to the corresponding "parent" attribute of the upper level: for example, weight $\mathrm{W}(T_{ijk})$ quantifies the contribution of basic "leaf" sub-attribute T_{ijk} (of Level 4) in the representation and definition of sub-attribute T_{ij} (of Level 3); instead, weight $W(M_l, T_{ijk})$ is the weight of the $l-th$ model with respect to the basic "leaf" sub-attribute T_{ijk}.

The weights $W(T_i)$, $W(T_{ij})$ and $W(T_{ijk})$ are calculated using pairwise comparison matrices: in particular, one pairwise comparison matrix is constructed for the attributes at the second level $S = 2$, one is constructed for each "set" of sub-attributes at level $S = 3$ that fall under the same "parent" attribute in the upper level $S = 2$, and one is constructed for each "set" of basic "leaf" attributes at level $S = 4$ that fall under the same "parent" sub-attribute in the upper level $S = 3$. The comparison matrix is a $(n \times n)$ square matrix, to be filled by experts, where *n* is the number of elements being compared. Attributes in each level are compared to each other with respect to their contribution in defining their "parent" attribute in the

upper level. For example, a (3×3) matrix is constructed to compare the basic sub-attributes $Q = T_{111}$, $Mp = T_{112}$ and $Dr = T_{113}$ (Level 4), with respect to their "parent" sub-attribute $D = T_{11}$ (Level 3). Typically, experts use a scale from 1 to 9 to evaluate the strength (i.e., the contribution) of each criteria with respect to the other; for example, the scale suggested by Saaty (2008) used to carry out a qualitative comparison between two attributes A and B, is the following:

1: A and B are equally important,

2: A is slightly more important than B,

3: A is moderately more important than B,

4: A is moderately-plus more important than B,

5: A is strongly more important than B,

6: A is strongly-plus more important than B,

7: A is very strongly more important than B,

9: A is extremely more important than B.

Another possibility is to define a scale of only the odd numbers between 1and 9 and use the even numbers to facilitate the judgment for intermediate situations (Zio 1996). See (Saaty 2008) and (Zio 1996) for more details.

A pairwise comparison matrix is made for each group of attributes in the same level (say, *s*) that falls under the same upper attribute in the upper level (s-1). The weight of each attribute is, then, determined by solving an eigenvector problem, where the normalized principal eigenvector provides the weights vector. For more details on how to calculate the weights of attributes, see (Saaty 2008); (Saaty & Vargas 2012); (Alexander 2012). Notice that the weights obtained should be normalized to sum to 1 as follows: $\sum_{i=1}^{n_T} W(T_i) = 1$, where n_T is the number of attributes under the "top" attribute T (i.e., model trustworthiness); $\sum_{j=1}^{n_{T_i}} W(T_{ij}) = 1$, where n_{T_i} is the number of sub-attributes under attribute T_i; $\sum_{k=1}^{n_{T_{ij}}} W(T_{ijk}) = 1$, where n_{T_i} is the number of basic "leaf" sub-attributes under sub-attribute T_{ij}.

For the tangible basic leaf sub-attributes T_{ijk}, a quantitative evaluation $T_{M_l,T_{ijk}}$ can be given by direct inspection and analysis of the models. Instead, if the basic leaf sub-attributes cannot be given a direct numerical evaluation (or if the analyst does not feel confident in carrying out this task), the scaling system explained above (i.e., scores from 1 to 9) can be adopted to provide a (semi-quantitative) relative evaluation of the leaf attributes T_{ijk} with respect to the risk models M_l available (guidelines are provided in Appendix A of this paper for relatively evaluating the basic leaf sub-attributes). The corresponding inter-level weights $W(M_l, T_{ijk})$ can, then, be obtained as $\frac{T_{M_l,T_{ijk}}}{\sum_{l=1}^{n} T_{M_l,T_{ijk}}}$. Note that the weights $W(M_l, T_{ijk})$ are thus normalized, i.e., $\sum_{l=1}^{\mathrm{n}} W(M_l, T_{ijk}) = 1$, where *n* is the number of models.

Finally, the normalized trustworthiness $T(M_l)$ of a model M_l is evaluated using a weighted average of the leaf attributes, as indicated in eq. (4):

$$T(M_l) = \sum_{i=1}^{n_{\mathrm{T}}} \sum_{j=1}^{n_{T_i}} \sum_{k=1}^{n_{T_{ij}}} W(T_i) * W(T_{ij}) * W(T_{ijk}) * \frac{T_{M_l,T_{ijk}}}{\sum_{l=1}^{n} T_{M_l,T_{ijk}}} \quad (1)$$

where $T_{M_l,T_{ijk}}$ is the numerical value that the basic "leaf" sub-attribute $T_{T_{ijk}}$ takes with respect to

model M_l, (for example, for attributes $Q = T_{111}$ variable $T_{M_l,T_{111}}$ equals the number of equations and correlations contained in M_l), n is the number of models to be compared, n_T, n_{T_i}, and $n_{T_{ij}}$ are defined above.

After obtaining the weight for each criterion with respect to the corresponding upper level criteria, a "global" weighting for each criterion with respect to the top goal T can also be obtained by multiplying its weight by the weights of its upper parent elements in each level: for example, the "global" weight of basic "leaf" sub-attribute T_{ijk} with respect to the "top" attribute (goal) T is given by $W(T_{ijk})$. $W(T_{ij})$. $W(T_i) = W_{global}(T_{ijk})$. For example, in the hierarchy tree Figure 1, the "global weighting" of the "consistency of data" (denoted by T_{221}) with respect to level of trustworthiness is obtained by multiplying its weight by the weight of quality of data (denoted by T_{22}) by the weight of strength of knowledge (denoted by T_2): $W(T_{221}).W(T_{22}).W(T_2) = W_{global}(T_{221})$. The trustworthiness $T(M_l)$ can then be expressed directly as a function of the "global" weights of the leaf attributes with respect to the top goal T:

$$T(M_l) = \sum_{i=1}^{n_T} \sum_{j=1}^{n_{T_i}} \sum_{k=1}^{n_{T_{ij}}} W_{global}(T_{ijk}) \frac{T_{M_l,T_{ijk}}}{\sum_{l=1}^{n} T_{M_l,T_{ijk}}} \quad (2)$$

Several considerations need to be made on the proposed approach. Clearly, there is no claim that the trustworthiness assessment method is comprehensive and complete. Attributes similar to those considered here have been already proposed and adopted in relevant works of literature: see, e.g., Flage & Aven (2009); Aven (2013b), where the strength of knowledge is assessed in terms of "phenomenological understanding", availability of reliable data", "agreement among peers" and "reasonability of assumptions", but there are other attributes that affect the level of trustworthiness as well.

In addition, the enumeration of some model leaf attributes (e.g., approximations, assumptions, formulas…) may be an "artifact" of presentation or interpretation, in absence of a protocol rigorously constructed to this aim. On the other hand, the following aspects should be considered. First, such a type of evaluation has been already used for evaluating some attributes in some relevant models e.g., evaluation of phenomenological understanding, availability of reliable data, reasonability of assumptions and agreement among peers, demonstrating the feasibility (Flage & Aven, 2009). Second, the issue of enumerating model assumptions and evaluating their quality have already been treated in several papers: see, e.g., (Aven, 2013b); ; (Boone *et al*., 2010). Then, most importantly, notice that the "direct enumeration" is not the only way to provide numerical values $T_{M_l,T_{ijk}}$ for the basic "leaf" attributes $T_{T_{ijk}}$with respect to the model M_l. As mentioned above, if the analyst does not feel confident in "counting" assumptions, formulas and correlations, he/she may resort to semi-quantitative scale (e.g., scores from 1 to 9), in order to provide a relative evaluation of a "leaf" attribute $T_{T_{ijk}}$with respect to the different risk models M_l's available (see for example the enumerating protocols in Appendix A, based on technical reports and experts' feedback).

4. Case study

In this section, the hierarchical tree-based framework is applied to a case study concerning the modeling of the residual heat removal (RHR) system of a nuclear power plant (NPP). In section 4.1, the system is described; in section 4.2, the characteristics of the two models used to represent the system (i.e.

the Fault Tree-FT and the Multi-States Physics-Based Model-MSPM) are presented in some detail; finally, in section 4.3, the proposed approach is applied to evaluate the trustworthiness of the two models.

4.1. The system

The Residual Heat Removal (RHR) system of a typical PWR reactor is taken as reference. The RHR is mainly used to remove the decay heat (residual power) from the reactor cooling system and fuel during and after the shutdown, as well as supplementing spent fuel pool cooling in the shutdown cooling mode for some types of reactors (NRC 2010). As illustrated in Figure 2, the main components of the RHR system are: pumps, heat exchangers, diaphragms, and valves. According to previous studies, it was found that 23% of RHR system failures are due to pumps failures, 58% are due to valves failures, while the rest of RHR system failures are due to other components' failures (Coudray & Mattei 1984).

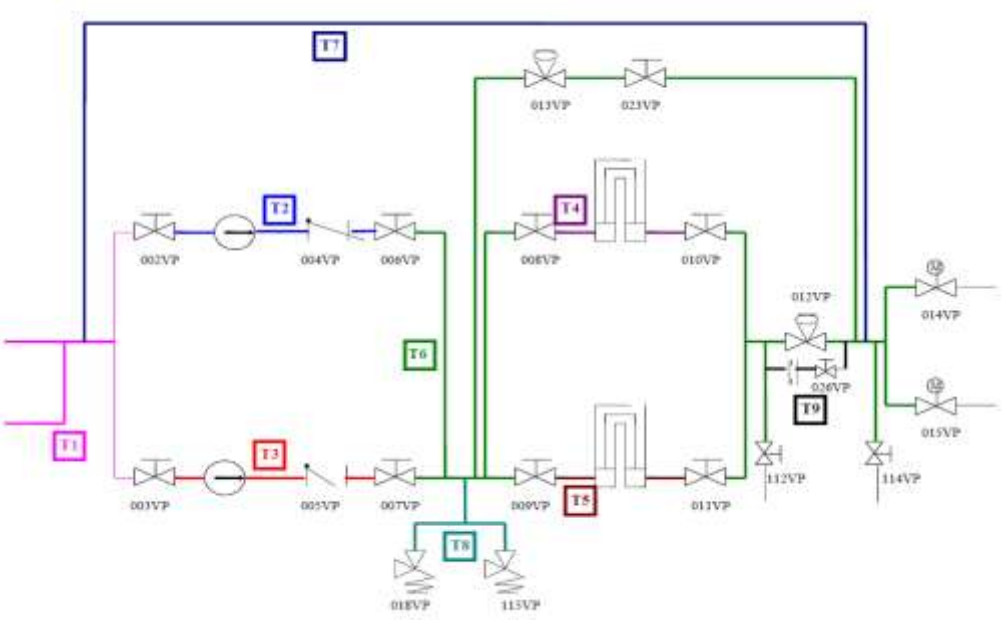


Figure 2 Schematic diagram of the RHR

4.2. Models considered

Two models have been considered for evaluating the reliability (resp., the failure probability) of the RHR system: a Fault Tree (FT) model (Section 4.2.1) and a Multi-State Physics-based Model (MSPM) (Section 4.2.2).

4.2.1. Fault Tree (FT) Model

The Andromeda software has been used for the analysis of the RHR's components failure modes and criticalities (importance analysis). The analysis is based on a logical framework for understanding the different possible ways in which the components and the system can fail. The failure probabilities used in the FT analysis are based on field experience feedback..

4.2.2. Multi-State Physics-based Model (MSPM)

Physics-based model (PBM) and multi-state model (MSM) are often used to describe the degradation processes of components and systems. Physics-based modeling aims to develop an integrated mechanistic description of the component/system life, consistent with the underlying degradation mechanisms (e.g. wear, stress corrosion, shocks, cracking, fatigue, etc.) by using physics knowledge and related

mathematical equations. Multi-state modeling is built on material science knowledge, degradation and/or failure data from historical collection or degradation tests, to describe the degradation processes in a discrete way (Gorjian *et al.,* 2010); (Di Maio *et al.*, 2015).

In general, MSM is able to describe the evolution of degradation in time, in terms of a range of states from "perfect functioning" to "complete failure". Since the degradation process is influenced by many factors, there are difficulties in estimating the transition rates required for the analysis of the degradation processes, especially for highly reliable components and systems (Di Maio *et al.*, 2015). It is also difficult to define precisely the states and the transitions between states in MSMs, due to the imprecise discretization of the degredation process and to data insufficiency (Lin *et al.*, 2015). Accordingly, a combination of the two models, namely the Multi-State Physics-based Model (MSPM), has been proposed, in which the state transition rate estimates are also based on physical models rather than operational data (Unwin *et al.*, 2011). Then, the whole process of transition and degradation can be described comprehensively by MSPM (Di Maio *et al.*, 2015).

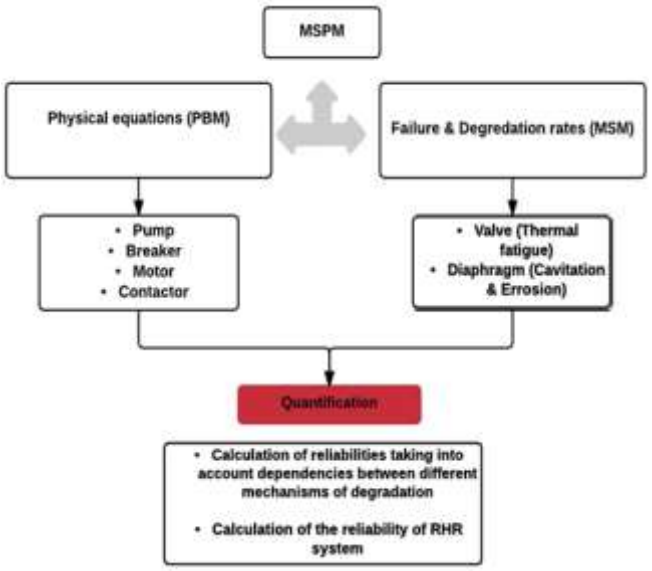


Figure 3 MSPM analysis: models of RHR components

In the present analysis of the case study, the main critical components were taken into account (i.e. pump, diaphragm, breaker, motor, contactor and valve). The MSM was used to model the pump, breaker, motor and contactor, while the PBM model was used to model the valve and diaphragm, taking into account the degradation dependency of the valve on the pump.

Figure 3 illustrates this setting. Three states were considered for the pump, including the fully functioning state, a degradation state corresponding to external leakage and the failure state. The breaker was modeled by a continuous-time homogeneous Markov model, taking into account the perfectly functioning and the failed states, and four types of failures were taken into account. Similarly a continuous-time homogeneous Markov model was developed for the analysis of the contactor and the motor, and four and two types of failures were taken into account for each, respectively.

On the other hand, the valve is subject to thermal fatigue that causes cracks or propagation of manufacturing defects, which are described by physical models and the related physical variables, such as

the coefficient of thermal expansion of the material, the modulus of elasticity, the Poisson ratio of the material, the elastoplastic strain concentration factors, the number of alternating cycles, etc. The crack initiation takes place when the amplitude of variation of the critical temperature ΔT_{lim} is exceeded, while the failure due to propagation of defects takes place when a specific number of cycles (operation demands) is exceeded. It should be noted that the total number of cycles executed over a period of time is calculated considering the degradation dependency of the valves on the degradation of the pump. In other words, when calculating the number of cycles executed by the valve, it is multiplied by a factor > 1 to consider the degradation dependency on the other components. Furthermore, the cavitation and the erosion are taken into account for analyzing the degradation and failure of the diaphragm. Different physical parameters are considered such as pressure, stress, dimension and other material-based characteristics. A threshold value at which the failure takes place is taken into account. More details about the system and the corresponding models cannot be reported here due to confidentiality reasons.

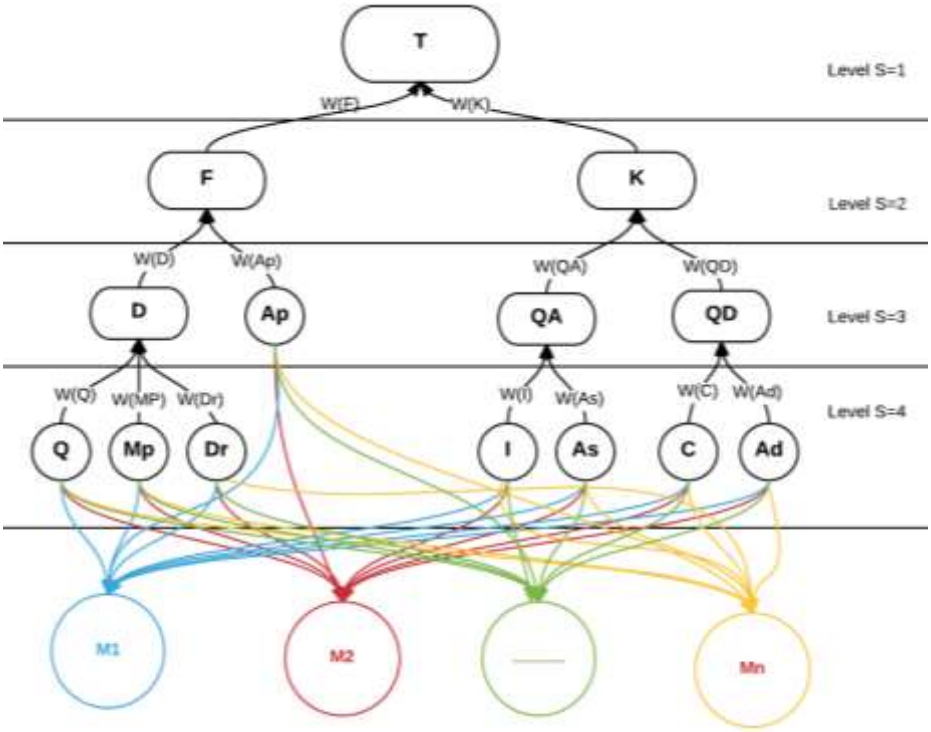
The results of MSPM and FT (using Andromeda software) are given in Table 3. The analysis shows similarities results in the first eight years. A difference between the two results starts to appear in the tenth year, showing a more rapid decline in the reliability values obtained by MSPM.

Table 3 Values of reliability

Time (years)		1	2	3	4	5	6	7	8	9	10
Reliability (FT)		0.779	0.607	0.473	0.368	0.288	0.224	0.175	0.136	0.107	0.083
Reliability (MSPM)		0.775	0.603	0.469	0.366	0.285	0.222	0.173	0.135	0.105	0.060

4.3. Evaluation of model trustworthiness

The analysis is carried out through two main steps: the first is an "upward" evaluation of the weight of each element in the hierarchy tree with respect to the top goal of model trustworthiness; the second is a



"downward" assessment of the model trustworthiness by means of a numerical evaluation of the basic "leaf" elements for both FT and MSPM models, as shown in Figure 4.

Figure 4 Hierarchical tree-based AHP model for the assessment of the trustworthiness of risk assessment models

With respect to the weights evaluation, experts were asked to fill the pairwise comparison matrices, in order to evaluate the importance of each attribute (criteria). As the experts were considered equally qualified, the weights obtained by solving the eigenvector problem of the pairwise comparison matrixes filled by the experts, were averaged. By way of example and only for illustration purposes, Table 4 shows a pairwise comparison matrix of the "leaf" sub-attributes $\mathrm{Q} = \mathrm{T}_{111}$, $\mathrm{Mp} = T_{112}$ and $\mathrm{Dr} = T_{113}$ of level s= 4. The attributes relative importances with respect to the parent attribute (level of detail) have been evaluated using the 1-9 scaling.

Table 4 Pairwise comparison matrix for "leaf" sub-attributes (Q, Mp and Dr) with respect to the "parent" D (level of detail)

	Q	Mp	Dr
Q	1	3	1
Mp	1/3	1	1/3
Dr	1	3	1

By solving the eigenvector problem for this matrix, we obtain the flowing weights: $W_{111} = 0.46$, $W_{112} = 0.21$, $W_{113} = 0.32$. Note that the weights of the three attributes in the example sum to one: $\sum_{k=1}^{3} W_{11k} = 1$.Table 5 shows the weighting factors obtained: in particular, the weights of each attribute with respect to the corresponding "upper level" parent (i.e., $W(T_i), W(T_{ij})\ and\ W(T_{ijk})$) as well as the "global" weight $W_{global}(T_{ijk})$ with respect to top goal T are given.

Table 5 Attributes weighting factors calculated using the AHP method

Parameter	Symbol	Level	Weight	Global weight
Model trustworthiness	T	S1	1.00	1.00
Modeling fidelity	F (T_1)	S2	0.35	0.35
Number of approximations	Ap (T_{12})	S3	0.54	0.19
Level of detail	D (T_{11})	S3	0.46	0.16
Number of equations and correlations	Q (T_{111})	S4	0.46	0.07
Number of model parameters	Mp (T_{112})	S4	0.21	0.03
Number of dependency relations	Dr (T_{113})	S4	0.32	0.05
Strength of knowledge	K (T_2)	S2	0.65	0.65
Quality of data	QD (T_{22})	S3	0.51	0.33
Amount of data	Ad (T_{221})	S4	0.60	0.20
Consistency of data	C (T_{222})	S4	0.40	0.13
Quality assumptions	QA (T_{21})	S3	0.49	0.32
Number of assumptions	As (T_{211})	S4	0.20	0.06
Impact of the assumptions	I (T_{212})	S4	0.80	0.25

The second step consists in an "upward" calculation, for the evaluation of the basic "leaf" attributes for each model. Actually, based on the data, information and knowledge available and used in the risk assessment analysis, two types of trustworthiness analysis have been implemented: one has been performed through a direct quantitative evaluation of the leaf attributes (e.g., for **Mp** (T_{112}) the number of model parameters are counted, for each model); the second is based on a semi-quantitative evaluation of the leaf attributes carried out through comparing the two models to each other and to the state of the art, and then, assigning a relative score (1-9) for each leaf attribute.

In order to do that, scaling guidelines have been defined based on several EDF's technical reports, (Burns 1980) and the feedback of experts, and scores of 1-9 have been defined (see Appendix A for details). Actually, we do not claim that those guidelines are complete and comprehensive, but they are sufficient for the context of the work. Relying on the guidelines of Appendix A, the data and technical reports used to perform the risk assessment, the relative score evaluation was performed for both FT and MSPM models: the results are reported in Appendixes B and C, respectively. In passing, notice that the evaluation of the attribute "Impact of the assumptions" ($I = \mathrm{T}_{212}$) is made as follows: a scale is given for each assumption and the scores are, then, averaged over all the assumptions.

On the basis of the relative scores selected, the trustworthiness evaluation was performed for both models, as illustrated in Table 6: the level of trustworthiness was found to be 0.4427 for Ft (M1) and 0.5573 for MSPM (M2).

We have applied the same method also to evaluate the models trustworthiness T using the direct quantification of the leaf attributes. The results are reported in Table 7. Table 8 shows all results.

Table 6 Comparison between FT and MSPM trustworthiness (relative scores)

Parameter	Symbol	Level	Weight	Global weight	Fault Tree		MSPM	
					Score	Weighted score	Score	Weighted score
Model trustworthiness	T	S1	1.00	1.00	-	4.65	-	5.85
Modeling fidelity	F (T_1)	S2	0.35	0.35	-	1.51	-	2.37
Number of approximations	Ap (T_{12})	S3	0.54	0.19	6	1.13	7	1.32
Level of detail	D (T_{11})	S3	0.46	0.16	-	0.38	-	1.04
Number of equations and correlations	Q (T_{111})	S4	0.46	0.07	3	0.22	8	0.60
Number of model parameters	Mp (T_{112})	S4	0.21	0.03	3	0.10	7	0.24
Number of dependency relations	Dr (T_{113})	S4	0.32	0.05	1	0.05	4	0.21
Strength of knowledge	K (T_2)	S2	0.65	0.65	-	3.14	-	3.49
Quality of data	QD (T_{22})	S3	0.51	0.33	-	2.06	-	2.25
Amount of data	Ad (T_{221})	S4	0.60	0.20	5	0.99	8	1.59
Consistency of data	C (T_{222})	S4	0.40	0.13	8	1.06	5	0.66
Quality assumptions	QA (T_{21})	S3	0.49	0.32	-	1.08	-	1.23
Number of assumptions	As (T_{211})	S4	0.20	0.06	5	0.32	6	0.38
Impact of the assumptions	I (T_{212})	S4	0.80	0.25	3	0.76	3.33	0.85

Table 7 Comparison between FT and MSPM trustworthiness (direct quantification)

Parameter	Symbol	Level	Weight	Global weight	Fault Tree		MSPM	
					Score	Weighted score	Score	Weighted score
Model trustworthiness	T	S1	1.00	1.00	-	58.45	-	113.59
Modeling fidelity	F (T_1)	S2	0.35	0.35	-	1.67	-	2.66
Number of approximations	Ap (T_{12})	S3	0.54	0.19	7	1.32	7	1.32
Level of detail	D (T_{11})	S3	0.46	0.16	-	0.35	-	1.34
Number of equations and correlations	Q (T_{111})	S4	0.46	0.07	1	0.07	9	0.67
Number of state rates and parameters	Mp (T_{112})	S4	0.21	0.03	8	0.27	18	0.61
Number of dependency relations	Dr (T_{113})	S4	0.32	0.05	0	0.00	1	0.05
Strength of knowledge	K (T_2)	S2	0.65	0.65	-	56.78	-	110.93
Quality of data	QD (T_{22})	S3	0.51	0.33	-	55.76	-	109.89
Amount of data	Ad (T_{221})	S4	0.60	0.20	275	54.70	549.15	109.23
Consistency of data	C (T_{222})	S4	0.40	0.13	8	1.06	5	0.66
Quality assumptions	QA (T_{21})	S3	0.49	0.32	-	1.02	-	1.04
Number of assumptions	As (T_{211})	S4	0.20	0.06	4	0.25	3	0.19
Impact (Sensitivity analysis)	I (T_{212})	S4	0.80	0.25	3	0.76	3.33	0.85

Table 8 Summary of models trustworthiness using relative scores and direct measures

	Fault Tree	MSPM
Normalized Trustworthiness (relative scores measures (1-9))	0.44	0.56
Normalized Model Trustworthiness (direct measures)	0.34	0.66

5. Discussion and Conclusion

In this work, we have developed a hierarchical tree-based decision making framework to assess the relative trustworthiness of risk models. The approach is based on the identification of specific attributes that are believed to affect the trustworthiness of the model. This is obtained through a hierarchical-tree based "decomposition" of the model trustworthiness into sub-attributes. The AHP method has been used to perform a weighted aggregation of the attributes to evaluate the model trustworthiness. The method has been applied to a case study involving the Residual Heat Removal (RHR) system of a Nuclear Power Plant (NPP). Two models of different complexity (i.e., FT and MSPM) have been considered to evaluate the system reliability and the trustworthiness of such models has been compared.

FT trustworthiness has been found to score 4. 65 out of 9, whereas MSPM has scored 5.85 or 0.34 and

0.66, respectively, by normalized direct measures of “leaf” attributes. The two results confirm the expectation that MSPM provides more trustworthy risk estimates than FT, due to the fact that it takes into account components failure dependency relations and time dependency of the degradation affecting the component.

Clearly, there is no claim that the trustworthiness assessment approach proposed is comprehensive and complete, as there exist other factors that affect the level of trustworthiness, which were not considered here. The method was, rather, a first attempt to systematically evaluate the models’ relative trustworthiness. Obviously, it impossible to remove completely subjectivity and expert judgment is still present, the method provided is an attempt to cast such expert judgment in a systematic and structured framework. Also, further studies should be performed to define the scaling guidelines for attributes evaluation and study how to integrate the level of trustworthiness in RIDM.

References

1. Alexander, M., 2012. Decision-Making using the Analytic Hierarchy Process (AHP) and SAS/ IML. *The United States Social Security Administration Baltimore*, pp.1–12.
2. Aven, T., 2013a. A conceptual framework for linking risk and the elements of the data-information-knowledge-wisdom (DIKW) hierarchy. *Reliability Engineering and System Safety*, 111, pp.30–36.
3. Aven, T., 2013b. Practical implications of the new risk perspectives. *Reliability Engineering and System Safety*, 115, pp.136–145.
4. Aven, T., 2016. Risk assessment and risk management: Review of recent advances on their foundation. *European Journal of Operational Research*, 253(1), pp.1–13. Available at: http://www.sciencedirect.com/science/article/pii/S0377221715011479.
5. Aven, T. & Heide, B., 2009. Reliability and validity of risk analysis. *Reliability Engineering & System Safety*, 94(11), pp.1862–1868.
6. Aven, T. & Zio, E., 2013. Model output uncertainty in risk assessment. *International Journal of Performability Engineering*, 9(5), pp.475–486.
7. Bani-mustafa, T. et al., 2018. Strength of Knowledge Assessment for Risk Informed Decision Making. In *Esrel*. Trondheim.
8. Berner, C. & Flage, R., 2016. Strengthening quantitative risk assessments by systematic treatment of uncertain assumptions. *Reliability Engineering and System Safety*, 151, pp.46–59.
9. Bjerga, T., Aven, T. & Zio, E., 2014. An illustration of the use of an approach for treating model uncertainties in risk assessment. *Reliability Engineering and System Safety*, 125, pp.46–53.
10. Boone, I. et al., 2010. A method to evaluate the quality of assumptions in quantitative microbial risk assessment. *Journal of Risk Research*, 13(3), pp.337–352. Available at: http://www.scopus.com/inward/record.url?eid=2-s2.0-77951165131&partnerID=40&md5=12a3caae6ff5f3fae9967becb6b35f17.
11. Burns, R.D., 1980. Wash 1400—Reactor safety study. *Progress in Nuclear Energy*, 6(1–3), pp.117119–117140.
12. Coudray, R. & Mattei, J.M., 1984. System reliability: An example of nuclear reactor system analysis. *Reliability Engineering*, 7(2), pp.89–121.
13. Cox, T. & Lowrie, K., 2015. Special Issue: Foundations of Risk Analysis.
14. Danielsson, J. et al., 2016. Model risk of risk models. *Journal of Financial Stability*, 23, pp.79–91.
15. Dezfuli, H. et al., 2010. NASA Risk-Informed Decision Making Handbook.
16. Droguett, E.L. & Mosleh, A., 2008. Bayesian methodology for model uncertainty using model performance data. *Risk Analysis*, 28(5), pp.1457–1476.
17. Eiser, J. et al., 2012. Risk interpretation and action: A conceptual framework for responses to natural hazards. *International Journal of Disaster Risk Reduction*, 1(1), pp.5–16.
18. EPRI, 2015. *An Approach to Risk Aggregation for RisK-Informed Decision-Making*, Palo Alto, California.
19. EPRI, 2012. *Practical Guidance on the Use of Probabilistic Risk Assessment in Risk-Informed Applications with a Focus on the treatment of Uncertanity*, Palo Alto, California.
20. Flage, R. & Aven, T., 2015. Emerging risk – Conceptual definition and a relation to black swan type of

events. *Reliability Engineering & System Safety*, 144(August), pp.61–67. Available at: http://www.sciencedirect.com/science/article/pii/S0951832015001982.

21. Flage, R. & Aven, T., 2009. Expressing and communicating uncertainty in relation to quantitative risk analysis. *Reliability: Theory & Applications*, 4(2–1 (13)).
22. Goerlandt, F. & Montewka, J., 2014. Expressing and communicating uncertainty and bias in relation to Quantitative Risk Analysis. *Safety and Reliability: Methodology and Applications*, 2(13), pp.1691–1699. Available at: http://www.crcnetbase.com/doi/abs/10.1201/b17399-230.
23. Gorjian, N. et al., 2010. A review on degradation models in reliability analysis. In D. Kiritsis et al., eds. *Engineering Asset Lifecycle Management: Proceedings of the 4th World Congress on Engineering Asset Management (WCEAM 2009), 28-30 September 2009*. London: Springer London, pp. 369–384. Available at: http://dx.doi.org/10.1007/978-0-85729-320-6_42.
24. Herbsleb, J. et al., 1997. Software quality and the capability maturity model. *Communications of the ACM*, 40(6), pp.30–40.
25. IAEA, 2006. *Determining the Quality of Probabilistic Safety Assessment (PSA) for Applications in Nuclear Power Plants*, Vienna: INTERNATIONAL ATOMIC ENERGY AGENCY. Available at: http://www-pub.iaea.org/books/IAEABooks/7546/Determining-the-Quality-of-Probabilistic-Safety-Assessment-PSA-for-Applications-in-Nuclear-Power-Plants.
26. INSAG, 2011. *A Framework for an Integrated Risk Informed Decision Making Process*, Vienna: INTERNATIONAL ATOMIC ENERGY AGENCY. Available at: http://www-pub.iaea.org/books/IAEABooks/8577/A-Framework-for-an-Integrated-Risk-Informed-Decision-Making-Process.
27. Lin, Y.-H., Li, Y.-F. & Zio, E., 2015. Fuzzy reliability assessment of systems with multiple-dependent competing degradation processes. *IEEE Transactions on Fuzzy Systems*, 23(5), pp.1428–1438.
28. Lin, Y., 2016. A holistic framework of degradation modeling for reliability analysis and maintenance optimization of nuclear safety systems.
29. Lin, Y.H. et al., 2013. Multi-State Physics Model for the Reliability Assessment of a Component under Degradation Processes and Random Shocks Multi-State Physics Model for the Reliability Assessment of a Component under Degradation Processes and Random Shocks.
30. Lopez Droguett, E. & Mosleh, A., 2014. Bayesian Treatment of Model Uncertainty for Partially Applicable Models. *Risk Analysis*, 34(2), pp.252–270. Available at: http://doi.wiley.com/10.1111/risa.12121 [Accessed June 26, 2017].
31. Di Maio, F., Colli, D., et al., 2015. A multi-state physics modeling approach for the reliability assessment of nuclear power plants piping systems. *Annals of Nuclear Energy*, 80, pp.151–165.
32. Di Maio, F., Turati, P. & Zio, E., 2015. Prediction capability assessment of data-driven prognostic methods for railway applications. In *Proceedings of the third European conference of the prognostic and health management society*.
33. Nasa, 2013. STANDARD FOR MODELS AND SIMULATIONS-NASA-STD-7009. , (I), pp.7–11.
34. Nicolas Zweibaum & Jean-Pierre Sursock, 2014. *Addressing multi-hazards risk aggregation for nuclear power plantsthrough response surface and risk visualization*, Palo Alto, California.
35. NRC, 2010. *Reactor Coolant System and Connected Systems*, Washington: NRC.

36. Oberkampf, W.L., Pilch, M. & Trucano, T.G., 2007. Predictive capability maturity model for computational modeling and simulation. *cfwebprod.sandia.gov*. Available at: https://cfwebprod.sandia.gov/cfdocs/CCIM/docs/Oberkampf-Pilch-Trucano-SAND2007-5948.pdf%5Cnfile:///Users/markchilenski/Documents/Papers/2007/cfwebprod.sandia.gov%0A/Oberkampf/cfwebprod.sandia.gov%0A 2007 Oberkampf.pdf%5Cnpapers://31a1b09a-25a9-4e20-879d-4.
37. Paté-Cornell, M.E., 1996. Uncertainties in risk analysis: Six levels of treatment. *Reliability Engineering & System Safety*, 54(2), pp.95–111.
38. Paulk, M.C. et al., 1993. Capability Maturity Model for Software, Version 1.1. *Software, IEEE*, 98(February), pp.1–26. Available at: http://www.sei.cmu.edu/library/abstracts/reports/93tr024.cfm.
39. Saaty, T.L., 2008. Decision making with the analytic hierarchy process. *International Journal of Services Sciences*, 1(1), p.83.
40. Saaty, T.L. & Vargas, L.G., 2012. *Models, methods, concepts & applications of the analytic hierarchy process*, Springer Science & Business Media.
41. Simola, K. & Pulkkinen, U., 2004. *Risk Informed Decision Making A Pre-Study*, Finland: Nordisk Kernesikkerhedsforskning.
42. Triantaphyllou, E. & Shu, B., 1998. Multi-criteria decision making: an operations research approach. *Encyclopedia of Electrical and Electronics Engineering*, 15, pp.175–186. Available at: http://univ.nazemi.ir/mcdm/Multi-Criteria Decision Making.pdf.
43. Unwin SD, PP Lowry, RF Layton, Jr, P.H.A.M.T., 2011. Multi-State Physics Models of Aging Passive Components in Probabilistic Risk Assessment. In *In International Topical Meeting on Probabilistic Safety Assessment and Analysis*. Wilmington, North Carolina: Amercian Nuclear Society, La Grange Park, IL., p. vol. 1, pp. 161–172.
44. Veland, H. & Aven, T., 2015. Improving the risk assessments of critical operations to better reflect uncertainties and the unforeseen. *Safety Science*, 79, pp.206–212. Available at: http://www.sciencedirect.com/science/article/pii/S092575351500154X.
45. Xu, L. & Yang, J.-B., 2001. *Introduction to multi-criteria decision making and the evidential reasoning approach*, Manchester School of Management.
46. Zeng, Z. et al., 2016. A hierarchical decision-making framework for the assessment of the prediction capability of prognostic methods. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 231(1), pp.36–52. Available at: http://dx.doi.org/10.1177/1748006X16683321.
47. Zio, E., 1996. On the use of the analytic hierarchy process in the aggregation of expert judgments. *Reliability Engineering and System Safety*, 53(2), pp.127–138.
48. Zio, E., Cantarella, M. & Cammi, A., 2003. The analytic hierarchy process as a systematic approach to the identification of important parameters for the reliability assessment of passive systems. *Nuclear Engineering and Design*, 226(3), pp.311–336.

Appendix A: Method used to translate the hierarchical tree attributes into a semi-quantitative scale

The following table presents the guidelines adopted in this paper to translate the attributes of the hierarchical tree into a semi-quantitative scale. Such guidelines are defined based on discussions and suggestions provided by EDF analysts, with relevant experience in the problem ad case study at hand.

Table A.1 A semi-quantitative scale for the hierarchical tree attributes

Parameter	**Translation "real number → scale 1/9"**
Number of approximations	Low number of approximation and low believed effect of their aggregate on the outputs: 9 few approximations with low effect of their aggregate: 7 moderate number of approximations with acceptable effect of their effect on the outputs: 5 high number of approximations with high effect of their aggregate on the outputs: 3 High number of approximations with sever effect of their aggregate on the outputs: 1 The even number are left for the intermediate cases
Number of equations and correlations	1-2 equations : 1 3 equations : 2 4 equations or 1 (Boolean logic equation) : 3 5 equations : 4 6 equations : 5 7 equations : 6 8 equations : 7 9 equations : 8 >9 equations : 9
Number of state rates and model parameters	0-2: 1 3-5: 2 6-8: 3 9-11: 4 12-14: 5 15-17: 6 18-20: 7 21-23: 8 >32: 9
Number of dependency relations considered	0 dependency relations considered : 1 1%-12.5% of the failures rates are considered dependent on the failure of other components: 2 13.5%-25%: 3

	26%-37.5%: 4 38.5%-50%: 5 51%-62.5%: 6 63.5%-75%: 7 76%-88.5%: 8 >88.5% All components failures are dependent on other components failures : 9
Consistency of data	The expert should give a score between 1-9 evaluating of the consistency of data, taking into account the source of data, its compatibility and relevance to the components that need to be analyzed. As in the case study the data is collected from the same type of reactors 900 Mwe, it is highly consistent: the consistency is given a score of 8. However, we cannot guarantee a perfect consistency, as the information about a specific component might be collected from other components that are similar but slightly different: e.g., the failure rate of RHR pumps is calculated taking into account failures of all pumps in the reactor.
Amount of data (Number/amount of sources)	The following classification is adopted according to the suggestions of EDF experts: > 25 reactor years of experience : 1 25-50: 2 51-100: 3 101-175: 4 176-275: 5 276-400: 6 401-550: 7 551-725: 8 Over 725: 9
Number of assumptions	Directly related to the actual number of assumptions used.
Impact (Sensitivity analysis and indications)	The impact is related to the assumptions. The difference between the values of failure rate with and without the assumption should be estimated. A score between 1-9 is given for each assumption, and the final score is then averaged over all assumptions. 1. No repairs: assuming no component repairs, at time 500, we obtain a probability of failure which is 500 times higher as compared to the case when the repair is considered (Figs 9-12 (Lin, 2016)) 2. One directional dependency: assuming only one-direction dependency of the valve degradation from the degradation and vibration of the pump,

	decreases the valve reliability of about 3 times (Figs 9-21(Lin, 2016)) 3. Human error: In case of human error (omission in closing the manual valve), we obtain a probability of failure of RHR which is 1.096 times higher. Nevertheless, the human error probability is very small. 4. No random shocks: assuming no random shocks results in a relative difference in the failure rate of the components. in particular, there is a reduction of (-2.99%-19823.08%) with respect to the case with the random shocks (Table II (Lin, 2016))

Appendix B: Trustworthiness attributes evaluation for Fault Tree (FT) M1

Table B.1 Trustworthiness attributes evaluation for Fault Tree (FT)

Parameter	**Direct score**	**Relative score**	**Note**
Number of approximations	7	6	7 minimal cut sets
Number of equations and Number of correlations	1	3	1 equation (Boolean logic): failure probability based on "rare event" approximation
Number of model parameters	8	3	8 failure rates for 8 basic events
Number of dependency relations	0	1	No dependency relations considered
Amount of data (Number/amount of sources)	275	5	EDF internal reports on data collected between 1980 and 1992, or 275 years reactor for each component.
Consistency of data	8	8	The data are collected from application of SAFO (OMF-reliability-centered-maintenance-feedback computer assisted collection on 7 CP1-CP2 sites and report on data. As this data is collected from the same type of reactors 900 MWe it is highly consistent. On the other hand, we cannot guarantee a "perfect" consistency, as the information about a specific component might be collected from other, similar but possibly different, components: e.g., the failure rate of RHR motor operated valves is calculated taking into account failures of all motor operated valves in the reactor.
Number of assumptions	4	5	1. No repairs 2. No dependency relations between components and failure mechanisms 3. Human error 4. No random shocks
Impact of the assumptions (average of the impact of the different assumptions considered)	3	Avg: 3 3	Based on the sensitivity analysis performed by (Lin, 2016) and the analysis performed using Risk Spectrum Software by EDF 1. No repairs: assuming no component repairs, at time 500, we obtain a probability of failure which

			is 500 times higher as compared to the case when the repair is considered (Figs 9-12 (Lin, 2016)) 2. No directional relation considered 3. Human error: In case of human error (omission in closing the manual valve) we obtain a probability of failure of RHR which is 1.096 times higher. Nevertheless, the human error probability is very small. 4. No random shocks: assuming no random shocks results in a relative difference in the failure rate of the components. in particular, there is a reduction of (-2.99%-19823.08%) with respect to the case with the random shocks (Table II (Lin, 2016))
		4 4 1	

Appendix C: Trustworthiness attributes evaluation for Multi-State Physics-based Model (MSMP) M2

Table C.1 Trustworthiness attributes evaluation for Multi-State Physics-based Model (MSMP)

Parameter	**Direct score**	**Relative score**	**Note**
Number of approximations	7	7	No relevant approximation
Number of equations and Number of correlations	9	8	4 multi-state models 3 physical equations for valve and diaphragm behavior 2 threshold equations for D_v and D_D (denote respectively: the number of cycles of solicitation of the valve over time and the thickness loss of the pipe over time)
Number of model parameters	18	7	-5 transitions rates in the multi-state model - 11 parameters for physical equations for the valve and diaphragm - 2 parameters for the modeling of number of cycles and thickness loss (18 parameters in total)
Number of dependency relations	1	4	1 dependency relation considered between the valve and the pump
Amount of data	549.15	8	-Pump : 621.95 years reactor -Breaker: 420 Years reactor -Contactor : 528.21 years reactor - Motor : 626.42 years reactor
Consistency of data	5	5	The data are collected from internal technical reports: -Pump 621.95 years reactor (PWR 900 MWe, PWR 1300 MWe, PWR N4) PWR 900: 2 PWR 1300, N4: 2 -Breaker 420 Years reactor (PWR1300 MWe, CPY) CPY: 18 PWR 1300:19 -Contactor 528.21 years reactor (1300 MWe, CPY, PWR N4) CPY: 26 PWR 1300: 48 PWR N4-1400: 29

			- Motor 626.42 years reactor (900 MWe, 1300 MWe, Palier PWR N4) CPY: 43 PWR 1300: 36 PWR N4-1400: 34 Even though the data collected in EDF internal reports comes from different sources with different types of reactors, it is still consistent as the different components are very similar.
Number of assumptions	3	6	1. No repairs 2. 1 directional dependency: the dependency of the valve degradation on the pump degradation and vibration 3. No random shocks
Impact of the assumptions (average of the impact of the different assumptions considered)	3.3333	Avg: 10/3 3 6 1	Based on the sensitivity analysis performed by (Lin, 2016): 1. No repairs: assuming no component repairs, at time 500, we obtain a probability of failure which is 500 times higher as compared to the case when the repair is considered (figs 9-12 (Lin, 2016)) 2. One directional dependency: assuming only one direction dependency of the valve degradation on the degradation and vibration of the pump decreases the valve reliability of about 3 times (Figs 9-21 (Lin, 2016)) 3. No random shocks: assuming no random shocks results in a relative difference in the failure rate of the components. in particular, there is a reduction of (-2.99%-19823.08%) with respect to the case with the random shocks (Table II (Lin, 2016))

Appendix II (P2): A Multi-Hazards Risk Aggregation Considering Maturity Levels of Risk Analysis

A Multi-Hazards Risk Aggregation Considering Maturity Levels of Risk Analysis

Tasneem Bani-Mustafa [(1)], Zhiguo Zeng [(1)], Enrico Zio [(1) (2)], Dominique Vasseur [(4)]

[(1)] *Chair on System Science and the Energetic Challenge, EDF Foundation*
Laboratoire Genie Industriel, CentraleSupélec, Université Paris-Saclay,
3 Rue Joliot Curie, 91190 Gif-sur-Yvette, France

[(2)] *Laboratoire, MINES ParisTech,* 1 Rue Claude Daunesse,, 06904 Sophia Antipolis*, France*

[(3)] *Energy Department, Politecnico di Milano, Via Giuseppe La Masa 34, Milan, 20156, Italy*

[(4)] *EDF R&D, PERICLES (Performance et prévention des Risques Industriels du parc par la simulation et les Etudes) EDF Lab Paris Saclay - 7 Bd Gaspard Monge, 91120 Palaiseau, France*

Abstract

Multi-Hazards Risk Aggregation (MHRA) aggregates risk over different risk contributors and provides a final risk index that permits the comparison with safety guidelines to support Decision Making (DM). The risk contributors assessment are conditional on many factors e.g., background knowledge, conservatism, sensitivity that are believed to determine the level of maturity of analysis and hence, realism of risk contributors indexes. Aggregation of risk contributor's values that are not identical in their degrees maturity and realism would lead to mathematically inconsistent and physically meaningless result that misinform the decision making. Hence, the difference in maturity, and the sources of heterogeneity that cause such differences, should be taken into account for supporting a reliable and accurate representation of risk in respect of DM.

In this paper, we propose a hierarchical framework to evaluate the level of maturity of risk contributors in the light of DM. The framework consists of four attributes that are believed to affect greatly the level of maturity of risk analysis i.e., uncertainty, conservatism, knowledge and sensitivity that are believed to affect the level of realism in the assessment of risk contributors. The knowledge attribute is in turn, broken down into five further sub-attributes i.e., availability of data, consistency of data, data reliability, experience, and value ladenness. Analytical Hierarchy Process (AHP) is adopted in this paper for the application of the framework and assessing the level of maturity. Reduced-Order Model technique is used to enable the application of the framework on real world complex problems. Then, the maturity level is integrated in MHRA by developing a two-dimensional risk aggregation method. Scoring protocols for evaluating the attribute were prepared to simplify the application of the framework and to reduce the subjectivity of the assessors. Finally, a numerical case study for the MHRA of a Nuclear Power Plant (NPP) is carried out to show the applicability and the plausibility of the methods. Please note that it is out of the context of this paper to show in details how to employ the maturity index in the process of DM.

Keywords

Probabilistic Risk Assessment (PRA), Risk Informed Decision Making (RIDM), Multi-Hazards Risk Aggregation (MHRA), Strength of Knowledge (SoK), Level of Conservatism, Uncertainty, Sensitivity Analysis, Nuclear Power Plant (NPP), Reduced Order Models.

1. Introduction

Risk can be defined as the possible harm that might occur to human or environment, and it needs to be considered in terms of both magnitude of detriment and its likelihood (INSAG 2011). In risk assessment we perform quantitative and qualitative measures of risk to ensure that it is maintained under the allowed safety limit. Risk assessment is based mainly on conceptual frameworks and qualitative assessment of risk that represents different systems and processes. The conceptual frameworks are in turn, built on a set of assumptions that are translated into quantitative assessments through representing them in mathematical forms, to provide measures and predictions of safety performance (Bjerga *et al.,* 2014); (NRC 2010); (Eiser *et al.*, 2012).

Recently there has been a great focus on risk and the developing a conceptual framework of risk as it is believed that risk interpretation play a vital role in Decision Making (DM) and therefor disasters reduction (Eiser et al. 2012). Actually, it is believed that in order to control and reduce risk, a comprehensive understanding of risk and the context of DM is required (Eiser et al. 2012). However, having a comprehensive understanding of risk requires knowing the risk, understanding it and having the ability to acknowledge it to help the decision maker to comprehend it (Simola & Pulkkinen 2004). Moreover, experts emphasize that relying solely on the numerical values of PRA as input value can be misleading for DM, as it does not capture or the important aspects related to DM (EPRI 2015).

As an example of how risk assessment is performed, the safety of French nuclear reactors is essentially based on a deterministic approach, supplemented by the Probabilistic Safety Assessment (PSA). PSA has been widely applied in various industries, e.g., nuclear, aerospace, defense, etc. Moreover, in 1995, NRC recommended in its final policy statement to increase the use of PSA in nuclear regulatory activities to the extent supported by the state of the art (NRC, 1995). A PSA is a systematic conceptual and mathematical tool that evaluate risks associated with a complex engineering systems such as Nuclear Power Plants (NPP) to support DM (Karanki *et al.*, 2009) and even more, the robustness of this decision is now, a matter of the quality of PSA (IAEA 2006).

PSA provides an overall quantitative and qualitative view of safety including both equipment and operators behavior by mainly: (i) identifying accidental scenarios leading to undesired consequences; (ii) assessing the probability of occurrence of these scenarios (Duménigo et al. 2008). Usually, different hazard groups (classification of hazard by its nature) are involved in a PSA (e.g., PSA of nuclear power plants usually involves hazard groups like fire, internal flooding, etc.). However, to make risk-informed decisions based on the results of PSA, Multi-Hazard Risk Aggregation (MHRA) is required: all relevant information on risk from different contributors is combined, arriving at an integrated risk index (EPRI 2015). Usually, risk-informed decisions are made by comparing the integrated risk index (e.g., core damage frequency, large early release frequency, risk increase, etc.) to safety goals and quantitative acceptance criteria.

Currently, most MHRAs are conducted by a simple arithmetic summing over the individual risk indexes for different hazard groups (EPRI 2015). For example, in current PSAs for Nuclear Power Plants (NPPs) in France, an overall risk index is computed by summing over the risk indexes of hazard groups like internal events, fire, external flooding, etc., which, permits comparing the overall risk index to safety goals and acceptance guidelines for Risk-Informed Decision Making (RIDM) (EPRI 2015). A main

criticism for the summation-based MHRA method is that it ignores the heterogeneities in the nature of the hazard groups, the degree of realism and the trust we have on the knowledge possessed over each one. Take again the PSA of nuclear power plants as an example. Among the hazard groups, the PSA model has been developed for internal events for many years, while relatively recently, the PSA for hazard groups like external flooding has started to be investigated (EPRI 2015). Therefore, we have more trust that the PSA for internal events is more realistic than for external flooding. Also, through the operation of US NPP, fire has been considered as a great contributor to the total risk, which might be due to the importance of the fire risk or/and due to the fact that it is characterized as immature and less realistic compared to some other initiating events; such as the internal events (Siu et al. 2015). The different levels of realism, which result from the difference in knowledge that supports the risk assessments, must be taken into account as they affect the risk-informed decisions based on the results of risk analyses (Aven 2013b).

Other sources of heterogeneity is the level of conservatism of the models, which, is based on the origin of the initiating events (EPRI 2015). In particular, for external hazards, due to the lack of data (testing, physical models, etc.), conservative assumptions are made regarding the impact of the hazards on the installation (EPRI 2015). Similarly, for the evaluation of the frequency of these hazard, studied at extreme levels of intensity, it is often difficult to establish a result in which we can have a great confidence. Additionally, many other key aspects are believed to influence the process of risk aggregation and RIDM such as: the interpretation of uncertainty and sensitivity analysis, value ladenness of the analysts and decision makers (), different natures of explicit and implicit knowledge, level of details and sophistication of risk analysis, etc., (EPRI 2012); (Zweibaum & Sursock, 2014). Actually, the aggregation over these hazards groups without considering the sources of heterogeneity and levels of realism and hence, trust that we possess for each hazard group, leads to a mathematically inconsistent and physically meaningless result (EPRI 2015). Nevertheless, these challenging aspects require drawing more attention on developing new PSA-supporting tool that allows pragmatically addressing them in order to help in risk-informed decision making (RIDM), especially, that the risk analysis cannot lead to a decision without the decision maker's judgment that reflects his subjectivity and preferences (Paté-Cornell, 1996).

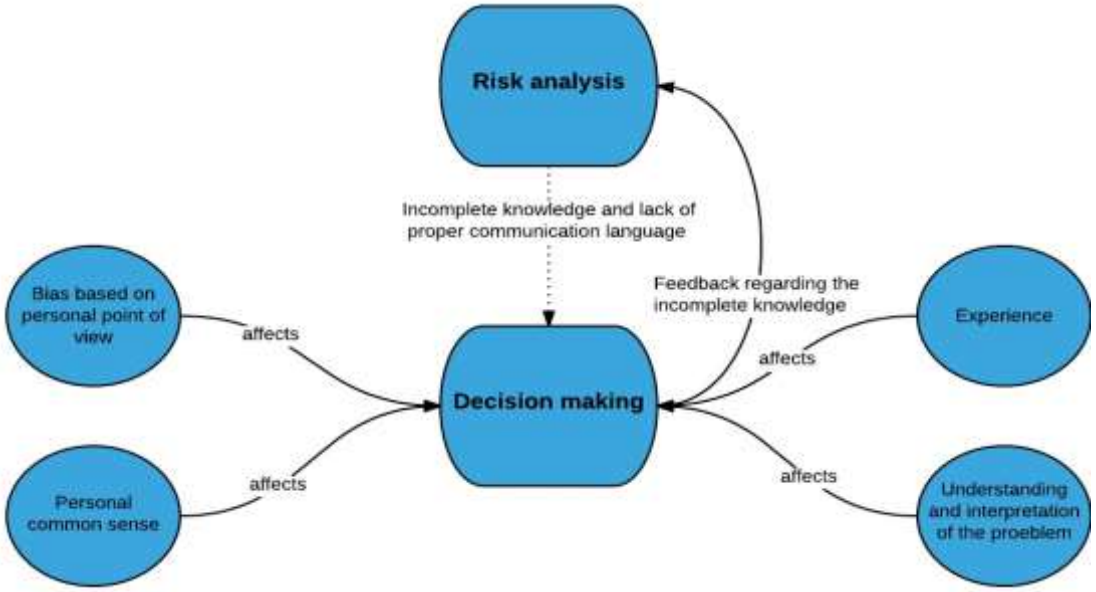


Figure 1 Risk informed decision making process, factors influencing decision making, and RIDM weaknesses

In this paper, we develop a MHRA supporting tool that considers the heterogeneities in the different contributors leading to different degrees of realism. The realism on a risk model is evaluated based on the concept of maturity. Maturity of a PSA is defined in this paper as the degree to which a PSA is correctly implemented in a way to reflect the available knowledge. The rest of this paper is organized as follows. In Sect. 2, we present a hierarchical framework for assessing the maturity of PSA and we develop evaluation (scoring) protocols to facilitate the process of assessment. Then, in Sect. 3, we develop an MHRA method that considers the maturity of the PSAs for different hazard groups. Section 4 applies the developed methods on a numerical case study. Finally, in Sect 5, we give a conclusion on the paper and we discuss the potential future work.

2. A hierarchical framework for PRA maturity assessment

In this section we discuss the different factor that are believed to affect the level of maturity of probabilistic risk analysis. In sect 2.1 we discuss the importance of introducing an index to evaluate the level of maturity and we mention some the factor that are introduced in the literature and believed to affect the level of maturity of risk analysis. In Sect 2.2 we propose four attributes for evaluating the level of maturity and we demonstrate their effect on the maturity and propose scoring protocols for the evaluation of the attributes.

2.1. Framework development

As illusrtated previously, many factors are believed to affect the the suitability of risk definition and risk aggergation. Emphasis is paid in the literature on importance of communicating these factor for better informing decision making (Flage & Aven 2009); (EPRI 2012); (Aven 2013b); (EPRI 2015); (Veland & Aven 2015). Some of these factors are: (i) background knowledge; (ii) level of uncertainty; (iii) level of conservatism; (iv) importance measures; (v) level of details and sophistication of the analysis; (vi) accuracy and precision in the estimation of the values of the model's parameters; (vii) level of sensitivity; (viii), and others (IAEA, 2006); (Flage & Aven 2009); (EPRI 2012); (Aven 2013a); (Aven 2013b); (Bjerga et al. 2014); (EPRI 2015); (Veland & Aven 2015); (Aven 2016); (Berner & Flage 2016).

In particular, MHRA includes aggregating risk from different contributor that have different degrees of realism, which in turn result from differences in characterizations e.g., of uncertainty, background knowledge, conservatism, etc. (EPRI 2015). Hence, MHRA needs to account for the these characterization and the different degrees of "realism" in the analysis of each risk contributors, (IAEA, 2006); (EPRI 2012); (EPRI 2015). Otherwise, the aggregation process would be mathematically inconsistent and physically meaningless results that misinform DM (EPRI 2015).

In this paper, we focus on communicating the factors that affects the degrees of realism in risk contributors, though a metric referred to as "level of maturity". The level Maturity of a PRA is expresses in this paper, the degree to which PRA is correctly implemented in a way that makes best use of the available knowledge to best represent the reality. In this section, we review the most relevant elements for mature risk assessment of PSA from literature, and develop a hierarchical framework for maturity assessment based on these elements.

2.2. Attributes elicitation and evaluation

In this section, four elements i.e., uncertainty, conservatism, knowledge and sensitivity (IAEA, 2006); (Flage & Aven 2009); (EPRI 2012); (EPRI 2015); (Aven 2016); (Berner & Flage 2016) relevant to the level of maturity and Risk-Informed Decision-Making (RIDM) are reviewed and discussed. In this review, we argue the importance of these attributes in determining the level of realism of probabilistic risk analysis and we propose evaluation protocols that are based on solid argument presented in the same sections. The overall hierarchical representation of the framework is illustrated in Figure 2.

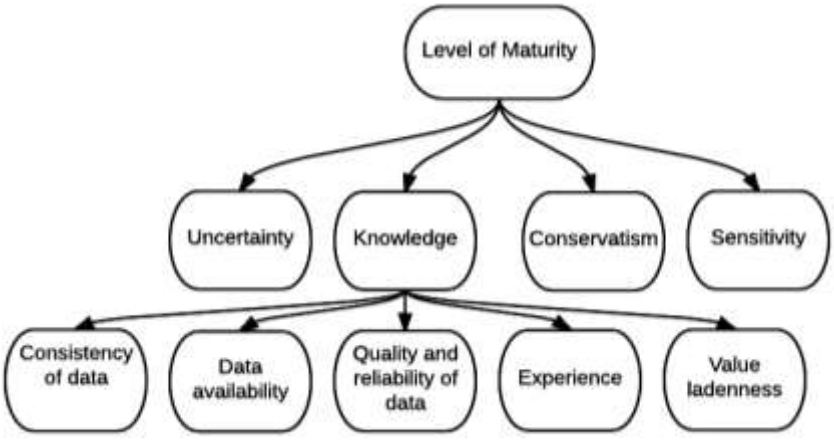


Figure 2 Level of maturity framework

2.2.1 Uncertainty

Uncertainty is defined as the imperfection of knowledge on the real value of a variable or its variability (Riesch 2013). Uncertainty is an important source of differences between the reality and the model predications (Bjerga et al., 2014). Hence, uncertainty affects greatly the credibility of PRA (Ferdous et al. 2013), (Abdo et al. 2017). This means that it reflects directly the level of maturity of the PRA and it should be addressed in its evaluation.

2.2.1.1 Uncertainty classification

Uncertainty can be classified relatively into different levels, depending on the degree of knowledge imperfection (Walker et al. 2003). For example, Wynne (1992) distinguishes four types of uncertainties depending on the level of knowledge: *"Risk"* where the system behavior is well known and quantifiable; *"uncertainty"* where the system parameters are known but the probability distributions are unknown; *"ignorance"* where the unknowns are unknown and finally; *"indeterminacy"* which underlies the indeterminacy in scientific knowledge construction with link to the tacit social knowledge. Walker et al. (2003) suggests three dimensions for uncertainty classification for uncertainty-based decision support purposes: the *"location"* where the uncertainty manifests itself within the model complexity, the *"level"* of uncertainty, which is, demonstrated by a spectrum between deterministic knowledge and absolute ignorance and finally, the *"nature"* of uncertainty which illustrates the type of uncertainty (epistemic or aleatory). The level of uncertainty is, further, classified into five progressive levels: determinism, statistical uncertainty, scenario uncertainty, recognized ignorance and total ignorance (Walker et al. 2003). Spiegelhalter and Riesch (2011) identify, within the spirit of Walker *et al.* (2003), five progressive levels of uncertainty for model-based risk analysis, each corresponds to a score are presented in Table 1.

Table 1 Uncertainty levels descriptions and scores with respect to the level of maturity

Level	Description	Score
Level 1 (uncertainty about the outcome)	This level of uncertainty manifests itself when the model and the parameters are known, and the analysis predicts a certain outcome with a probability *P* (e.g., the uncertainty about the outcome in most traditional mathematical and philosophical problems of probability theory)	5
Level 2 (uncertainty about the parameters	The model is known but its parameters are not. If the parameters are known then the model would predict an outcome with probability *P* and exhibit an uncertainty of level one. This type of uncertainty arises due to lack of empirical information on the model parameters (e.g. input parameters related to Large Break in Primary Circuit of a Nuclear Power Plant that has never occurred)	4
Level 3 (uncertainty about the model)	It reflects the likelihood of the competing models' abilities to reflect reality. This type of uncertainty is due to the model structure itself and the computer implementation of the model (Walker et al. 2003)	3
Level 4 (uncertainty about the acknowledged limitations and	This level covers any known limitations in understanding and modelling abilities, which arises from the inevitable assumptions and simplifications made such as: data extrapolations, limitation in the computations, and any aspects that we are aware that they have been omitted.	2

implicit assumptions-unmodeled uncertainty)		
Level 5 (Uncertainty about unknown inadequacies)	It is the unrecognized uncertainty or as it was referred to by Donald Rumsfeld the "unknown unknowns", which corresponds to the unforeseen events, unmodeled and unmodlable uncertainty. This type of uncertainty are usually acknowledged by brainstorming of the possible scenarios, or by the introduction of what so called 'fudge factors'.	1

Whilst this classification seems to be too crude and simple to be correct, it satisfactorily covers, at least from this problems' perspectives, the three dimensions defined by Walker *et al*. (2003) i.e., *"location"*, *"level" and "nature"* of uncertainty. For example, the definition of Level 1 of uncertainty, refers to the aleatoric nature of uncertainty, while Levels 2-5 cover the epistemic nature of uncertainty. Also, where the five levels vary progressively from the known to the unknown-unknown, they simultaneously refer to its location i.e., parameter, model and context of uncertainty. Moreover, the applicability and handleability of this method makes it a better choice to serve the context of this work.

2.2.2 Conservatism of analysis

Conservatism in PRA refers to desire of cautiousness by overestimating the risk. The conservatism in PRA arises from different considerations and perspectives such as the concerns regarding the lack of knowledge about the nature and magnitude of the hazard (Viscusi et al., 1997). This leads to the implementation of the concept of "Better safe than sorry", Samuel Lover, which is further translated to the preference of overestimating the risk rather than underestimating it. For example, selecting risk estimate that exceeds the mean of value of the probability distribution at the 95^{th} percentile, which, means that there is a 95% probability that the risk is over estimated and 5% is underestimated (Perhac Jr 1996).

Although the conservatism is usually anticipated to increase safety, some counter-arguments still exist on its influence on safety margin (Perhac Jr 1996). It has been argued that conservatism cannot be advised only from a risk-aversion point of view, and that the cumulative effects of conservatism on decision-making, regulations and risk management are unacceptable (Perhac Jr 1996), (Viscusi *et al.*, 1997). In particular, the effect of conservatism is not taken into account from a firm empirical sense (Viscusi *et al.*, 1997), which might be, in some contexts, perceptive for the analysts by giving a false assurance of safety, leading to worst consequences of risk (Whipple 1987). In fact, the overall effect of conservatism on safety (whether that conservatism is protective or not), depends greatly on the assumptions made, and the context of decision making (Whipple 1987).

Viscusi *et al.* (1997) argue that though conservative risk estimates increases the risk magnitude, the implications of this increase on the safety is still a matter of the decision-makers' actions. They have showed through a cost-benefit based study (number of lives saved per unit cost) that unlike conservative assessment the mean parameter approach would result in enhanced judgment policies that would enhance

the safety. This can be explained by the shift of prioritization of decision maker. Moreover, recent studies conclude and explicitly recommend that conservatism should be avoided in the light of some decision making contexts like: comparing options and studying the effects of potential risk reducing measures (Aven 2016). The degree of conservatism should be complied with the decision contexts and requirements of the PRA. Otherwise, it might reduce the maturity level and sometimes mislead the decision maker.

2.2.2.1 Conservatism classification

All of the arguments mentioned in the previous section, lead to question how to classify of levels of conservatism in the light of the maturity and its consequences on safety. At a first glance, classifying the levels of conservatism depending on the level of knowledge seems plausible, especially that conservatism represents a practical act performed to deal with uncertainties and lack of knowledge. However, this is not valid considering its implication on safety, where other aspects should be taken into account aside from strength of knowledge e.g., the context of decision making. Aven (2016), highlights the conservatism in risk analysis as a multi-dimensional concept, reinforcing the former arguments of experts about the real effect on safety (mentioned in Sect. 2.2). This is done by firstly addressing the meaning of conservatism, secondly relating it to the strength of knowledge and thirdly evaluating its usefulness in the context of decision-making. In this vision, he compares conservative risk indexes (i.e., based on conservative assumptions) to three cases: (I) risk indexes based on best estimate assumptions; (II) risk indexes based on true value parameters (III) risk indexes based on true value parameters with a defined confidence statement. Then, for these cases (I-III), he defines the possible states of knowledge on which the assumptions or risk parameters are based and finally, the possible contexts of decision, and tries to relate it to the consequences on safety (Aven 2016). Hereafter, we extend the work of Aven (2016) and define three main types of risk index estimates: (i) best judgment estimates (based on best judgment of assumptions and parameters); (ii) true value with a high confidence (based on strong knowledge); (iii) true value with a low confidence (based on weak knowledge). Then, for two context of decision making, i.e., comparing alternatives and comparing the risk indexes to acceptance limit, we compare the three defined estimate types (i-iii) to the conservative estimates (based on conservative assumptions) and give scores for each possible scenario with respect to level of maturity and safety. In other words, we are comparing the estimates that are based on assumptions chosen to be conservative (for cautiousness reasons) to those estimates that are based on the best judgment or true values of assumptions and parameters. Figure 3-5 illustrate the different score for each corresponding scenario.

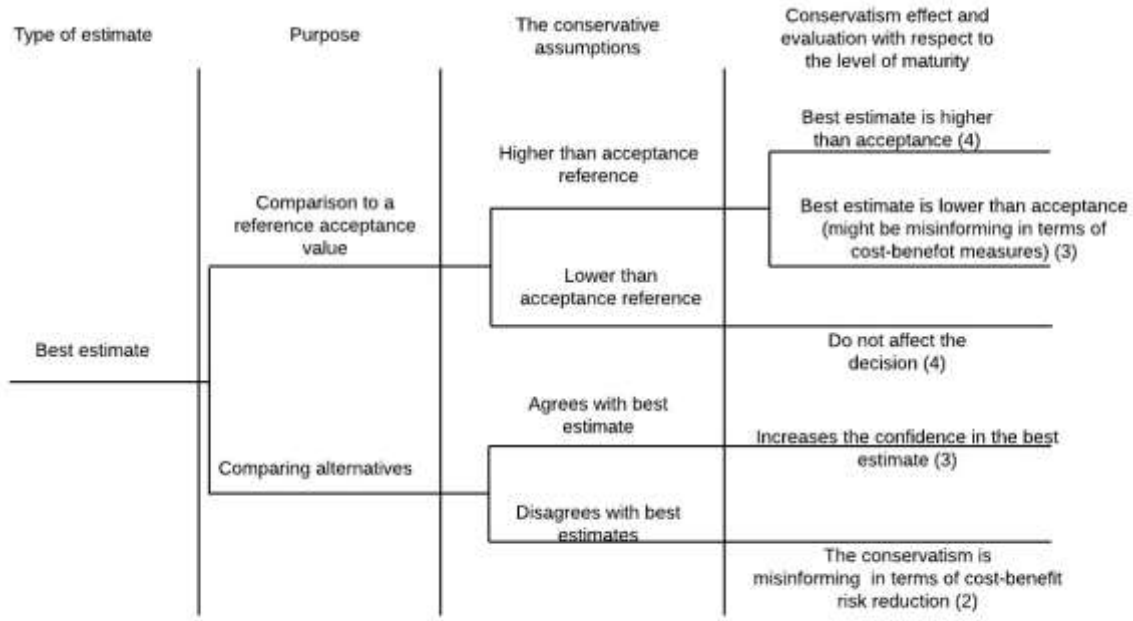


Figure 3 Evaluation of the conservatism in the light of level of maturity (conservatism VS Best estimate)

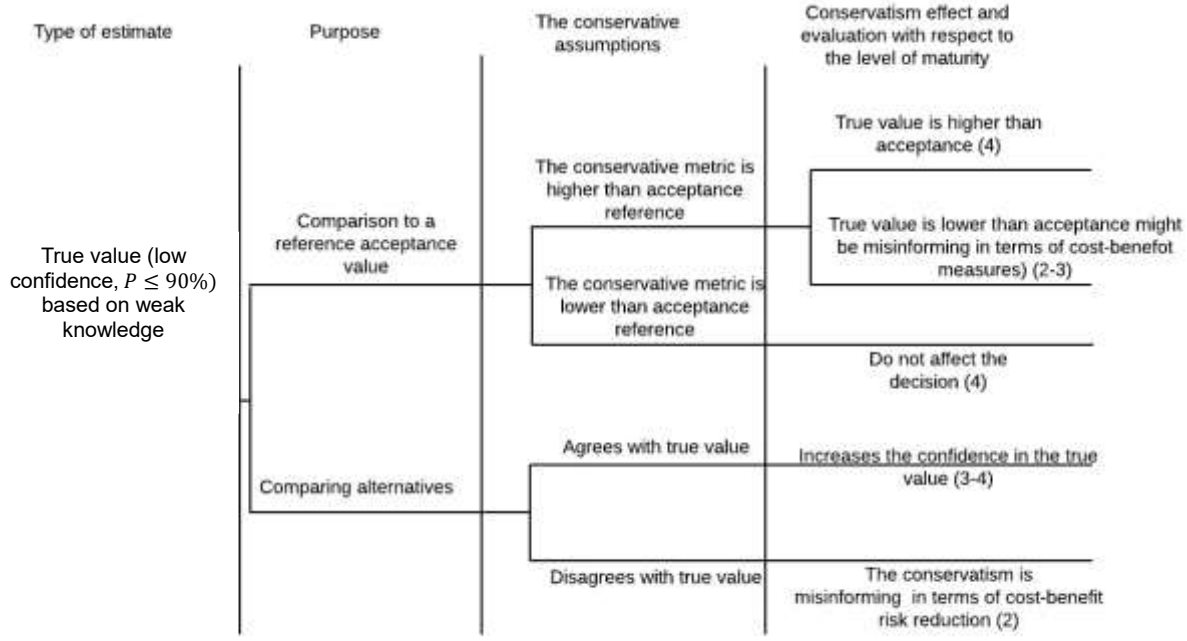


Figure 4 Evaluation of the conservatism in the light of level of maturity (conservatism VS True value/weak knowledge)

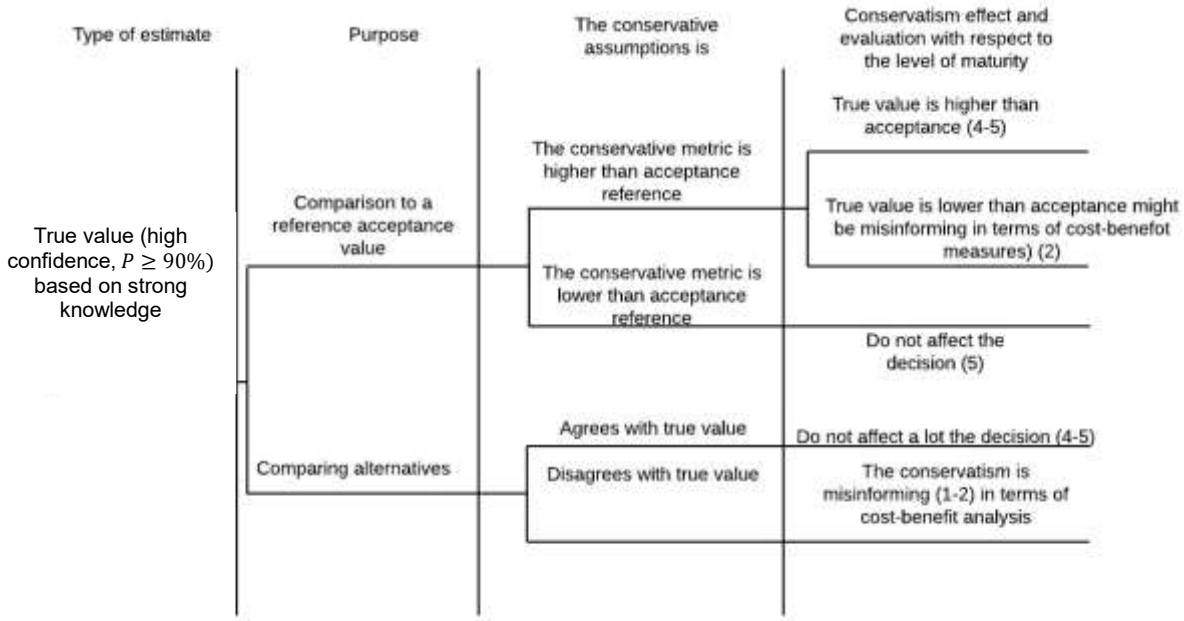


Figure 5 Evaluation of the conservatism in the light of level of maturity (conservatism VS True value/strong knowledge)

2.2.3 Knowledge

Knowledge is the second top tier of the four levels knowledge-hierarchy (DIKW hierarchy). It is the yield of a combination of data, information, experience and judgment to be used in decision-making (Aven 2013a). Knowledge manifests itself in three main forms; explicit and implicit, and tacit (Davies 2015).

It is said that “You can't manage what you can't measure." Peter Drucker. To best employ knowledge, one should be able to state its level. This led experts in safety and risk assessment to emphasize the importance of considering the background knowledge on which risk assessment is based, especially for Risk-Informed Decision-Making (RIDM) purposes (Aven 2013a), (Aven 2013b), (Aven & Krohn 2014), (Berner & Flage 2016), (Askeland et al., 2017), (Aven 2017), (Khorsandi & Aven 2017). This argument is visibly manifested in the new risk perspectives, which considers strength of knowledge in addition to the traditional elements i.e., scenarios, likelihood and consequences (Aven 2013b), (Aven & Krohn 2014), (Bjerga & Aven 2015), (Aven & Ylönen 2016). For these reasons, evaluating strength of knowledge should be considered in evaluating the models’ credibility and maturity.

2.2.3.1 Knowledge evaluation

Different attributes can be considered to evaluate the strength of knowledge such as, the amount of data and information, its suitability and usefulness, the human cognition regarding a specific phenomenon, the experience on the technology and of the analysts etc. There are however two main methods on which most of the strength of knowledge assessment approaches are based, in safety and risk assessment: a semi-quantitative approach for evaluating the Strength of knowledge (Goerlandt & Montewka 2014), and the assumption deviation risk by (Aven 2013b). In the earlier, the authors identify four main criteria for evaluating the strength of knowledge: the phenomenological understanding, the reasonability and realism of assumptions, the availability of reliable and relevant data and the agreement among peers (Goerlandt &

Montewka 2014). Based on the degree of fulfilling the criteria, the strength of knowledge is classified crudely to minor; moderate; and significant. The later method is based mainly on evaluating the criticality of the main assumptions on which probabilistic risk assessment is based, by evaluating the deviation from assumption, the uncertainty of this deviation and the strength of knowledge on which the former are based. Accordingly, the number of assumptions and the criticality of deviation from assumption, indicates the strength of knowledge on which the probabilistic risk assessment is based (Aven 2013b). However, one should not forget that in addition to the explicit properties of knowledge, it has also implicit and tacit properties (Davies 2015), and although it cannot be directly stated or documented, it contributes to the individual and organizational performance (Talisayon 2009). Obviously, in Flage and Aven (2009) the reasonability of assumptions and agreement among peers, are partially related to the implicit and tacit knowledge. However, this framework does not cover convincingly the assessment of tacit knowledge (e.g., agreeing on an assumption or assessment does not necessarily make it good), hence, the carriers of implicit and tacit knowledge (assessors) should rather be themselves evaluated.

In fact, several researches have emphasized on the importance of evaluating the value ladenness and confidence in experts' judgment. For example, Zio (1996) points to the fact that expert's judgment is subject to inevitable bias that lead experts that have the same background knowledge, to make different judgment, and he defines few attributes that are believed to affect the experts' judgment such as, the personal interest, the personal knowledge, the degree of independence, the experience etc. Other aspects such the situational limitations, choice space, agreement among peers and stake holders are included as well to assess the quality and robustness of assumptions on which, are made by the assessors, and the analysis are based (Boone et al., 2010), (Van Der Sluijs et al. 2005), (Kloprogge et al., 2011). Above all, one can argue that there are many other attributes that could be used to better represent the level of knowledge. However, Flage and Aven (2009) method in evaluating the strength of knowledge seems very plausible and relevant to the context of this problem except that it doesn't take into account the assessment of the experts who make the assumptions and the reasoning of the analysis, neither the availability of trustable predicting models. In this paper, we adjust and expand Flage's and Aven's (2009) method in Table 2, and add a new main attribute i.e., value ladenness of the assessor to the framework, to be adapt to the context of this paper.

Table 2 Level of knowledges' attributes evaluation guidelines

	Score	**1**	**3**	**5**
Data availability *(A)*	Amount of data/field data ($Sc_{3,1}$)	No data or the data are so limited and (can extracted only from the same type of NPPs)	The data are available and can be extracted from any other NPP	The data are Available in abundance (can be extracted easily from so many sources and places worldwide)
Data consistency *(Co)*	Source of data ($Sc_{3,2}$)	The data are extracted from other sources that is not related directly to the technology (not the exact same type of component)	Other NPPs of the same type and technology	Field data from the same power plant, and related to the same type of components

Quality and reliability of data *(Q)*	Quality of Data ($Sc_{3,3}$)	Based on experts elicitation	Data are calculated using statistical models	Data are both assumed and calculated using computer physical and mathematical models	Data are extracted using computer mathematical and physical models	The data are measured precisely and accurately, and then modeled
	Quality of assumptions ($Sc_{3,4}$)	Represents strong simplifications		Represents moderate simplifications	Represents reasonable simplifications	
Experience *(E)*	Phenomenological understanding ($Sc_{3,5}$)	The phenomena involved are not well understood		The phenomena involved are understood but not completely	The phenomena involved are very well understood	
	Experience and knowledge regarding the hazard group ($Sc_{3,6}$)	No experience at all		Experienced such an event in other industries	This event is quite common and we have a wide experience in	
	Availability of models ($Sc_{3,7}$)	Models are non-existent or known to give poor predictions.		The models used are believed to give predictions with moderate accuracy	The models used are known to give predictions with the required accuracy	
Value ladenness of the analysts *(VL)*	Agreement among peers ($Sc_{3,8}$)	There is strong disagreement among experts		There is slight agreement among experts	There is broad agreement among experts	
	Expert years in experience in the field and performance measure ($Sc_{3,9}$)	has quite short experience in risk assessment of NPPs		It is his specialty and he practiced through training courses regarding the same type of NPPs	Expert in this domain (long experience)	

2.2.4 Sensitivity

A mathematical model might embrace errors due to the lack of the knowledge regarding the input parameters or due the numerical methods used to solve the model (Cacuci *et al.*, 2003). The effects held by such errors are very important and need to be evaluated as it reflects the range of the trustworthiness and validity of the model. This is, done by sensitivity analysis (Cacuci *et al.*, 2003).

Sensitivity analysis is generally used to determine how a dependent variable can be changed and affected by the change of the input independent variable (Cacuci *et al.*, 2003). This is usually used to determine the critical control points and to prioritize additional data collection (Christopher Frey & Patil 2002). Moreover, it is implemented to provide the comprehensive understanding needed for a reliable use of the model, through highlighting and quantifying its most important features (Cacuci *et al.*, 2003), as well as verifying and validating it (Christopher Frey & Patil 2002).

In safety and risk assessment, sensitivity analysis can be useful in many ways. In particular, sensitivity analysis complements the risk analysis to inform decision-making (Borgonovo & Cillo 2017), where it helps to identify the uncertain inputs that contributes to the uncertainty in the outputs and consequently, affect the decision making process (Zio & Pedroni 2012). For example, in PRA of Nuclear Power Plants (NPPs), sensitivity analysis is required to study the impact of different model basic events' probabilities on the decision (Reinert & Apostolakis 2006). Also, the importance of an assumption in a risk prediction model can be evaluated through altering the input parameters or the background knowledge related to the given assumption, which helps in identifying the critical assumptions and the risk of their deviations (Goerlandt & Montewka 2014). Furthermore, sensitivity analysis is recommended in the practice of risk assessment to reduce -in some cases- the unnecessary conservatism (NRC 2011). From these perspectives, sensitivity analysis is considered an indispensable tool for evaluating model credibility and maturity.

2.2.4.1 Sensitivity evaluation

Flage and Aven (2009) suggested integrating the sensitivity concept as a main component of the uncertainty in order to have a holistic picture of the uncertainty beyond the concept of the probability. A rough semi-quantitative evaluation of sensitivity has been introduced with three levels of classification: significant sensitivity, moderate sensitivity and minor sensitivity. The simplicity of this method makes it very helpful in the context of decision making, as it gives an indication on the associated consequences and implications of parameters' deviations. On the other hand, it doesn't show how to apply the sensitivity analysis, neither how to translate it into a sensitivity level. For this reasons, we suggest to complement Flage and Aven (2009) by using a one-at-a-time index and then, converting it into a relative scores that represents the sensitivity levels suggested by Flage and Aven (2009).

In one-at-a-time method the sensitivity index S, measures the average of relative change in the dependent (output) variable $Y(x_i)$ by altering one input (x):

$$S = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y(x_{i+1})-Y(x_i)}{Y(x_i)}\right| \quad (1)$$

where x_i is the input parameter, n is the number of times that the analyst would apply the sensitivity measures by altering one input by an estimated suitable value e.g., ±20%, ±SD (standard deviation) (Hamby 1994) or ±4SD (Downing *et al.,* 1985). However, we are considering a ±50% altering parameter in this study to represent more clearly the sensitivity of parameters, as we are more concerned with PSA models that have a linear relation with the basic events (each basic event is unique and appears only one time in a given minimal cutset).

In this kind of analysis converging from (0) indicates the insensitivity of the model, while diverging from (0) indicates its sensitivity. After applying these analysis, the results need to be converted into discrete scores (e.g., 1: minor, 2: moderate, 3: significant (Goerlandt & Montewka 2014)) that indicate their levels. A sensitivity score (1-5) is assigned for the sensitivity index relying on the degree that the index converge or diverge from 0 as illustrated in **Error! Reference source not found.**.

Table 3 Scores representation of the sensitivity measure

Interval	*S:* ≤0.10	*S* : 0.10-0.25	*S*: 0.25-0.45	*S*: 0.45-0.70	*S*: ≥0.70

Level of sensitivity	1	2	3	4	5
Score	5	4	3	2	1

3. PRA maturity assessment

In this section we implement the developed framework through Analytical Hierarchy Process (AHP) method and we develop a method for evaluating the level trustworthiness of the overall risk analysis.

The evaluation process is carried out through two main steps. In the first step (Sect. 3.1), we evaluate the maturity attribute for each risk contributor on the required level i.e., the level of risk parameters, the level of hazard group etc. Then, we aggregate the maturity attributes scores for the overall hazard group. Finally in Sect 3.2, we aggregate the overall risk considering the levels of maturities of each hazard group.

3.1. Evaluation of the level of maturity for a single hazard group

For each criterion and sub-criterion defined in Figure 1, a semi-quantitative evaluation is carried out by assigning a relative score from 1 to 5, based on a set of pre-defined scoring criteria as illustrated in Sect. 2.2.1-2.2.4. The next step is to aggregate the scores of different attributes (criteria) to assess the overall maturity of a risk contributor. In this paper, the maturity level is calculated as a weighted average of the scores of the attributes.

$$m_i = \sum_{j=1}^{N_p} \sum_{i=1}^{n_d} w_i . w_{i,j} . Sc_{i,j} \quad (2)$$

where m_i is the level of maturity for the i-th hazard group that need to be evaluated, $w_{i,j}$, Sc_{ij} and w_i are respectively the weight and the score the j-th sub-attribute in the i-th attribute, and the weight of the i-th attribute. N_p is the total number of attributes and n_d *is* the number of sub-attributes related to the i-th evaluation criterion. The relative weight of each attribute w_i and sub-attribute $w_{i,j}$ should be evaluated. In this paper, we adopt Analytical heretical Process (AHP) as will be shown later in the case study. Where, pairwise comparison matrixes are developed for each group of daughter attributes (fall under the same parent attribute) to compare their relative importance in defining their parent attribute. Experts were asked to fill the constructed pairwise matrixes. A score of 1 was given to the equally important attributes, and a score of 5 was given when the first attribute is extremely more important than the other one. The weight of each attribute is, then, determined by solving an eigenvector problem, where the normalized principal eigenvector provides the weights vector. However, it is out of the context of this paper to show in details how to apply AHP method (for more information on AHP method see (Saaty 2008); (Saaty & Vargas, 2012)).

After constructing the AHP hierarchy and determining the relative weight of the attributes, Eq. 2 can be applied to determine the level of maturity. However, evaluating the level of maturity abstractly on the level of hazard group is not realistic, where PRAs of complex systems and their hazard groups embrace, often, multiple PRA elements that have different levels of maturity and need to be evaluated separately. In this light, we borrow in this work the idea of Bani-Mustafa *et al.* (2018), where the PRA model needs to be deconstructed into its constituting atomic elements. The PRA model is then reduced by taking into account

the most important atomic elements and then accounting to their contribution in building the model as the following (for more details, see (Bani-Mustafa *et al.* 2018)):

- Calculate the risk R_{O_i} for each operation state O_i
- Rank R_{O_i} in descending order
- From the descending-order list, find the number of operation states n_O that correspond to the amount of risk that needs to be assessed e.g., 80% of the risk
- At each operation state in the reduced order PRA model, calculate the risk $R_{O_i.S_i}$ for each scenario S_i
- Rank $R_{O_i.S_i}$ in descending order
- From the descending-order list, find the number of scenarios $n_{O,S}$ that correspond to the amount of risk that needs to be assessed e.g., 80% of the risk on this operation state
- At each operation state at each scenario in the reduced order PRA model, calculate the risk $R_{O_i.S_i,MCS_i}$, for minimal cutset MCS_i
- Rank $R_{O_i.S_i,MCS_i}$ in descending order
- From the descending-order list, find the number of minimal cutsets $n_{O,S,MCS}$ that correspond to the amount of risk that needs to be assessed e.g., 80% of the risk on this operation state
- At each minimal cutsets in the reduced-order PRA model, identify the related basic events BE_q
- Calculate the risk contribution of each scenario at a given operation state to the reduced-order overall risk.

Assuming that the risk on reduced-order model is expressed by elementary reduced-order models, which represent the risk for each scenario at a given operation state, the weight of each elementary risk model can be expressed by:

$$W_l = \frac{R_l}{\sum_{l=1}^{n_l} R_l} \tag{3}$$

where R_l is the risk of elementary reduced-order model and n_l is the number of elementary reduced-order models and expressed by $n_l = n_O \times n_S$.

- Calculate the weight $W_{l,q}$ of each basic event in a given elementary reduced-order model by:

$$W_{l,q} = \frac{I_{l,q}}{\sum_{q=1}^{n_{l,q}} I_{l,q}} \tag{4}$$

where $n_{l,q}$ is the number of basic events in the l-th elementary reduced-order model, $I_{l,q}$ is the Fussell-Vesely importance measures of the q-th basic event in the l-th elementary reduced-order model.

- Evaluate the maturity on each basic event by:

$$m_{l,q} = \sum_{i=1}^{N_p} \sum_{j=1}^{n_d} w_i . w_{i,j} . Sc_{i,j,l,q} \tag{5}$$

where $m_{l,q}$ is the level of maturity for the q-th basic event in the l-th elementary reduced-order model, $w_{i,j}$ and $Sc_{i,j,l,q}$ are respectively the weight and the score of the j-th sub-criterion in the i-th evaluation criteria for the q-th basic event in the l-th elementary reduced-order model.

- Evaluate the maturity m_i for the total hazard group by:

$$m_i = \sum_{l=1}^{n_l} \sum_{q=1}^{n_{l,q}} W_l \,.\, W_{l,q} . m_{l,q} \quad (6)$$

3.2. Risk aggregation considering maturity levels

In this paper, we adopt the perspectives of (Aven 2013b) that when characterizing risk, not only the probability index estimated by PRA, but also the knowledge that supports the PRA should be taken into account. Hence, in this paper, we use a tuple (R_i, m_i) to quantify the risk associated with hazard group i, where R_i and m_i are respectively the risk index and is the maturity level of the i-th hazard group PRA model, evaluated based on the method presented in Sect. 2.

A two-stage aggregation method is, then, developed for MHRA considering maturities of hazard groups. Suppose we have n_h hazard groups with the risk tuple $(R_i, m_i), i = 1,2, \cdots, n_h$. The overall risk can, then, be represented as a risk tuple (R, M) and computed in two steps:

Step 1: Aggregation of risk indexes. Risk indexes are aggregated following the summation rule:

$$R = \sum_{i=1}^{n_h} R_i \quad (7)$$

where R is the risk index after considering all the hazard groups. The physical meaning of R is the aggregated risk index, when we have complete confidence on each of the hazard group.

Step 2: Determine the maturity of the aggregated risk assessment:

In this paper we present two different possibility for aggregating and presenting the overall maturity of PRA model.

In the first suggestion, the maturity can be as well represented for the overall framework by applying a weighted average the maturities from each hazard group, considering the risk contribution for each hazard group:

$$M = \sum_{i=1}^{n_h} W_i . m_i = \sum_{i=1}^{n_h} \sum_{l=1}^{n_l} \sum_{q=1}^{n_{l,q}} W_i . W_l \,.\, W_{l,q} . m_{l,q} \quad (8)$$

where W_i is weight of the hazard group and calculated as the following:

$$W_i = \frac{R_i}{\sum_{i=1}^{n_h} R_i} \quad (9)$$

In the second suggestion, we borrow the aggregation idea from (Oberkampf *et al.,* 2007). The approach, recommends computing and presenting a set of three maturity scores. These scores consist of the minimum, average and maximum scores over all the hazard-group maturity-scores being aggregated, as the following:

$$M = \left[M_{min}, M_{avg}, M_{max}\right] = \left[\min_{i=1,2,\ldots n} m_i \,, \frac{1}{n_h} \sum_{i=1}^{n_h} m_i \,, \max_{i=1,2,\ldots n} m_i\right] \quad (10)$$

where M is the maturity triplet level of the PRA considering all the hazard groups, m_i is the maturity score of the i-th hazard group, n_h is the number of hazard group considered in the risk assessment model.

The aggregated risk, denoted by the quadruple $\left(R, M_{min}, M_{avg}, M_{max}\right)$, can, then, be used to support risk-informed decision making. Suppose we are considering the risk of a specific event. Instead of directly comparing R to the acceptance threshold, the maturity level should also be considered: when maturity level

is low, a larger safety margin is required; while when maturity level is high, a risk close to its threshold value might be accepted. The relationship between maturity level and the required safety margin should be determined, based on the severity of the consequence of the event.

Another possibility is, to represent the maturity as a vector of maturity attributes, which can be useful, as it allows the decision maker to know the weakness points in the analysis that leads to low maturity and ask the analyst to enhance the modeling and make further investigations if possible. The maturity of the hazard group is therefore, represented by:

$$m_i = (Sc_1, Sc_2, \dots Sc_{N_p}) \tag{11}$$

where Sc_{N_p} is calculated using a weighted average of the basic events in the reduced-order model of the given hazard group by:

$$Sc_{N_p} = \frac{1}{n}\sum_{l=1}^{n_l}\sum_{q=1}^{n_{l,q}}\sum_{j=1}^{n_d} W_l . W_{l,q} . Sc_{j,l,q} \tag{12}$$

where $Sc_{j,l,q}$ is the j-th sub-criteria score for the q-th basic event in the l elementary reduced-order model.

The maturity level M is represented by a vector of the scores average over all hazard group and calculated by:

$$M = (\frac{1}{n}\sum_{h=1}^{n} Sc_1 , \frac{1}{n}\sum_{h=1}^{n} Sc_2 , \dots \frac{1}{n}\sum_{h=1}^{n} Sc_n) \tag{13}$$

$$M = \begin{bmatrix} Sc_1 \\ Sc_2 \\ \vdots \\ Sc_{N_p} \end{bmatrix} = \frac{1}{n}\sum_{h=1}^{n} \begin{bmatrix} Sc_{1,h} \\ Sc_{2.h} \\ \vdots \\ Sc_{N_p,h} \end{bmatrix} \tag{14}$$

4. Case study

In this section, we apply the developed framework on a case study of two hazard groups in NPPs. The level of maturity assessment framework is, then, applied on the BEs and the total level of maturity for the overall hazard group is calculated by aggregating the BEs' maturities. The needed data and information that supports the model development were found in the technical reports provided by EDF, which are not mentioned here for confidentiality reasons.

4.1 Description of the hazard groups

In this case study, we consider two hazard groups PRAs, i.e., external flooding and internal events that were developed using Risk Spectrum Professional software by Electricité De France (EDF).

In PRA of NPP, "External flooding" refers to the overflow of water due to naturally induced external causes, e.g., tides, tsunamis, dam failures, etc. (IAEA, 2003).

"Internal events" refer to undesired events that might lead to loss of important components and consequently systems and that originate within the NPP itself (EPRI, 2015), such as structural failures in the components, safety systems operation errors, etc. (IAEA Safety Standards Series, 2009).

4.2 Evaluation of the level of maturity for external flooding hazard group

As illustrated in Sect. 3.2, the assessment needs to be carried out at the level of small risk

contributors. Hence, we first start by deconstructing the PRA model for each hazard group into their constituting atomic elements. The model is then, reduced to most important elements following the approach suggested in Bani-Mustafa et *al*. (2018).

Following the procedure in Sect 3.2, only one operation state i.e., $n_O = 1$ is found to cover more than 80% of the risk. Similarly, only one scenario $n_{1.S} = 1$ is found to cover more than 90% of the risk at this operation state.

At operation state O_1, and scenario $S_{1,1}$, five minimal cutsets $n_{1.1.MCS} = 5$ are found to cover more than 80% of this risk of $S_{1,1}$. Notice that the basic events are then, identified for the five corresponding minimal cutsets as presented in Table 4.

Table 4 Basic events included in the reduced-order model

Symbol	Basic event
BE1	External flooding with water level A inducing a loss of offsite power
BE2	Loss of auxiliary feedwater system due to the failure to close the isolating valve
BE3	Loss of component cooling system because of clogging
BE4	Failure of all pumps of the Auxiliary feedwater (AFW) system
BE5	Failure of the turbine of AFW system
BE6	Failure of the Diesel Generator A
BE7	Failure of the Diesel Generator B
BE8	Failure of the common diesel generator
BE9	Failure of pumps 1 and 2 of AFW system
BE10	Failure of pumps 2 and 3 of AFW system

The levels of maturity for the basic events in Table 4 need to be evaluated using the developed method in Sect 3.2. In the following, we illustrate in detail how to apply the developed framework on a basic event namely "External flooding with water level A inducing a loss of offsite power" (BE_1). For the other basic events, we directly give the results in Table 6..

As shown in Eq. (5), the level of maturity of a basic event is evaluated as a weighted average over the maturity attributes and sub-attributes illustrated in Figure 5. Hence, the weights of the maturity attributes and sub-attributes need to be determined. AHP method is adopted in this paper for this purpose (Saaty 2008). Two pairwise matrixes are constructed and filled by experts. The first is a 4×4 comparison matrix, constructed for evaluating the weights W_i (relative importance) of the attributes under level of maturity in defining their "parent" attribute i.e., level of maturity. The second is 5×5 comparison matrix constructed for comparing the weights $W_{i,j}$ (relative importance) of the strength of knowledge "daughter" attributes (i.e., sub-attributes under the strength of knowledge). For more illustration on AHP method and pairwise comparison matrixes see (Saaty 2008). The results are presented in Table 5. Notice that, the weights are evaluated only once and used for the evaluation of all the basic events.

The next step for evaluating the level of maturity is to assess the attributes and sub-attributes presented in Figure 1 for BE1 in the light of the guidelines presented in Sect. 2. In this basic event, the probability was calculated by extrapolating the probability distributions based on observed data to the

extreme water flowrate (i.e., flowrates that have never occurred). In more details, the following steps were performed:

- Height at which different events (failures of specific elements) take places where defined.
- The water flowrate was predicted for the given heights at the NPP platform ensuring to cover each flowrate that can lead to the given water height at the platform.
- The flowrate was multiplied by safety factors.
- The "return period" (the period on which you can have a flood with a given flowrate) were obtained by the same law that was used to estimate the millennial flooding flowrate of the river of interest.
- The return periods for flowrates of interest were then, calculated by extrapolating the flooding data curves toward extreme values (at low probabilities) of flow at the platform of the power plant.
- The frequencies (frequency =1/period de retour) were then, calculated rounded, mean values obtained by the law for the flowrates of the Millennial Flood.
- The frequency of each interval is chosen to be the maximum frequency at the whole height interval.
- No uncertainty analysis was taken into account for estimation the frequencies of the critical heights.
- Due the basin special characteristics, the analysts are forced to consider the "theory of renewal" (combining two statistical models of occurrence of events and their magnitude together).

Comments:

- Experts have confidence in the calculation used to convert the heights into flowrates because they are based on solid deterministic models.
- Experts have doubts on extrapolating the frequency to the extreme flowrates.
- This result is also to be considered with caution since they are based on the current limited models and knowledge.
- Multiplying the flowrates by safety and augmentation factors is considered conservative.
- The characteristics of the river basin are special in view of the evolution of the distributions of extreme floods, which opens more room for uncertainty.
- Used the renewal approach is considered conservative.
- High uncertainty is presented in the analysis.

From the previous arguments, one can notice that there is uncertainty about the acknowledged limitations and implicit assumptions (unmodeled uncertainty). This meets Level 4 of uncertainty, which leads the analysts to assign a score of (2) From Table 1.

For the conservatism attribute, it is not possible in this case to consider the conventional acceptance criteria (e.g., acceptable core meltdown of 10^{-4}) since we are considering only one hazard group.

Accordingly, experts were asked to assign an artificial value for the acceptable external flooding probability, in order to compare it to the estimated external flooding risk value of our model of interest. Now, since the analysis of the external flooding probability is based on hydrodynamic model then it is considered to be realistic but with low level of confidence. From figure 4, since we are comparing the risk metric to an acceptance criteria, it was found that the conservative estimates are misinforming. A score of 2 was assigned for the conservatism.

The sensitivity of this basic event is calculated by Eq.1. The basic events probability is altered by 50%. Which leads to the total change in the model output by 50% (since this basic event appears in each minimal cutset and has a Fussell-Vesely importance measure of 1). From Table 3, this corresponds to a level 4 of sensitivity, which in turn, corresponds to a score of 2 in the light of maturity.

The same way of reasoning was adopted for evaluating the scores of knowledge attributes. The results are shown in Table 5. The maturity attributes scores are then, aggregated by Eq. 5. The level of maturity for BE_1 is found to be 2.15.

Table 5 Assessment of level-3 knowledge "leaf" attributes (BE_1)

Attribute	U	C	S	K							
Sub-attribute	-	-	-	A	Co	QD	QA	Ph	Ex	AM	P	PM
W_i	0.30	0.15	0.15							0.40		
$W_{i,j}$	-	-	-	0.25	0.06	0.17	0.17	0.10	0.05	0.10	0.05	0.05
Score	2	2	1	1	5	3	2	3	5	3	5	5

The same steps are repeated for all the basic events and presented in Table 6. The final step before evaluating the overall level of maturity for external flooding hazard group $m_{ext-flood}$, is to determine the weights of each basic event, in a given elementary model and the corresponding elementary model by Eq.3 and Eq. 4.

From Eq. 3, the weight of the elementary model is: $W_1 = \frac{R_l}{\sum_{l=1}^{1} R_l} = 1$

From Eq. 4, the weight of the basic event in the given elementary model is: $W_{1,1} = \frac{I_{l,1}}{\sum_{q=1}^{n_{l,q}} I_{l,q}} = 0.320$

The same procedure are repeated for each basic event and the results are presented in Table 6. Finally, the overall level of maturity is evaluated by Eq. 6. The level of maturity is found to be $m_{ext-flood} = 2.45$.

Table 6 Knowledge assessment and aggregation over the basic events

BE	BE1	BE2	BE3	BE4	BE5	BE6	BE7	BE8	BE9	BE10
$m_{l,q}$	2.150	1.488	2.690	3.948	4.002	4.002	4.038	3.962	3.908	3.908
$I_{l,q}$	1.000	0.9020	0.553	0.182	0.141	0.127	0.121	0.045	0.028	0.028
$W_{l,q}$	0.320	0.289	0.177	0.058	0.045	0.041	0.039	0.014	0.009	0.009

$W_{l,q} \times m_{l,q}$	0.688	0.429	0.476	0.230	0.180	0.163	0.156	0.057	0.035	0.035

The same steps are repeated for the internal events hazard groups and the maturity was found to be $m_{internal} = 3.87$.

Finally, for risk maturity aggregation, we adopt the first technique presented in Sect. 3.3 where the risk is represented as a risk tuple (R, M). Please note that the risk presented here after are artificial and the real number that provided by EDF are not presented for some confidentiality reasons.

External flooding risk tuple: $\left(R_{ext-flood}, m_{ext-flood}\right) = (1.5^{-5}, 2.45)$

External flooding risk tuple: $(R_{internal}, m_{internal}) = (1.2^{-7}, 3.87)$

First, by Eq. 7 the total risk is calculated arithmetically $R = 1.512^{-5}$. Then the level of maturity is calculated by Eq. 8. Two variables need to be considered, the level of maturity m_i of a given hazard group, and its corresponding weight (relative importance). The hazard group weight is calculated by Eq.9 and found to be $W_{ext-flood} = 0.992$ and $W_{internal} = 0.008$. Finally, the overall maturity is found to be 2.462 and the risk tuple is $(1.512^{-5}, 2.45)$.

4.3 Results and discussion

As expected, the level of maturity for internal events ($m_{internal} = 3.87$) is higher than that for external flooding ($m_{ext-flood} = 2.45$). This mean that the analysis and the results of the internal events are more realistic than these for external flooding. This can be explained by the fact that unlike external flooding, risk analysis for internal events hazard group in NPP has been performed for all power plants all over the world, which in turn, created the opportunity to develop solidly the appropriate models, level of details and base knowledge required for realistic evaluations (EPRI 2015). This lead to a relatively well established highly mature PRAs (EPRI, 2012). On the other hand, as seen in the example above: most of the risk is contributed by BE_1, BE_2 and BE_3 (they have relatively high importance measures), which corresponds respectively to: (1) “external flooding with water level A inducing a loss of offsite power”; (2) “loss of auxiliary feedwater system due to the failure to close the isolating valve”; (3) “loss of component cooling system because of clogging”. The three basic events probabilities are obtained based on relatively, low level of knowledge, high misinforming conservatism and high uncertainty. For example, BE_1 the probability of this basic event is calculated by extrapolating the distributions based on observed data to the extreme water flowrate (i.e., flowrates that have never occurred) and that the probabilities of floods were taken as mean values without considering the uncertainty analysis. In addition, the characteristics of the river basin are special in view of the evolution of the distributions of extreme floods, which opens more room for uncertainty.

The overall risk is represented by $(R, M) = (1.512^{-5}, 2.45)$. Most of the risk and level of maturity in this tuple is on account of external flooding hazard group, which in turn, explains the low level of maturity on the overall risk.

5. Conclusions

In this paper, we have proposed a method for evaluating qualitatively the different degrees of realism and maturity in risk contributor's analysis. In this framework, we tried to focus on the attributes that are believed and emphasized in the literature to affect the level of realism and maturity of analysis, and most importantly, the process of decision making. The framework is based on four main attributes: uncertainty, conservatism, strength of knowledge and sensitivity. The strength of knowledge attribute, was further broken into five sub-attributes (data availability, data consistency, source of data, quality and reliability of data, experience and value ladenness of the analysts. Analytical Hierarchy Process (AHP) is adopted to apply the framework, where pairwise comparison matrixes were built to estimate the relative weights of the attributes. An assessment protocols were developed to facilitate the process of attributes evaluation for a given problem. In addition, the reduced order model approach in (Bani-mustafa et al. 2018) is adopted to evaluate the maturity on the level of constituting elements (basic events), which in turn, leads to a more relevant and accurate results. Finally, the developed framework was applied on two hazard groups in Nuclear Power Plants (NPP); namely, external flooding and internal events. The application of the framework on the case study, has showed its operability. The level of maturity of external flooding is $m_{ext-flood} = 2.45$ and for internal events $m_{internal} = 3.87$. The results of the application correspond to expectations, where the of internal events' PRAs practice is more well established and more mature compared to external flooding. The overall risk is found to be $(1.512^{-5}, 2.45)$. The low level of maturity for the overall risk is due to low maturity of external flooding that contributes highly to the overall risk. This in fact, emphasize the importance of accounting for the level of maturity of a given hazard group where it can be informing for the decision maker in contexts where an option needs to be chosen, or for doing further analysis to enhance the maturity before making a decision.

A potential limitation of the developed approach is the subjectivity of the analysts who are evaluating the relative importance (weights) as well as the scores of the maturity attributes. In addition, we do not pretend that the framework itself is complete in terms of the attributes and factors that affect the level of maturity. However, it still stands a good starting point for overcoming the heterogeneity in the maturity level of the hazards group that in turn lead to mathematical inconsistent and physically non-meaningful results. Finally, please note that it is out of the context of this paper to show in details the process of Decision Making (DM) given this maturity index.

References

1. Abdo, H., Flaus, J.M. & Masse, F., 2017. Uncertainty quantification in risk assessment-Representation, propagation and treatment approaches: Application to atmospheric dispersion modeling. *Journal of Loss Prevention in the Process Industries*, 49, pp.551–571.
2. Askeland, T., Flage, R. & Aven, T., 2017. Moving beyond probabilities ??? Strength of knowledge characterisations applied to security. *Reliability Engineering and System Safety*, 159(October 2016), pp.196–205.
3. Aven, T., 2013a. A conceptual framework for linking risk and the elements of the data-information-knowledge-wisdom (DIKW) hierarchy. *Reliability Engineering and System Safety*, 111, pp.30–36.
4. Aven, T., 2017. Improving risk characterisations in practical situations by highlighting knowledge aspects, with applications to risk matrices. *Reliability Engineering & System Safety*, 167, pp.42–48. Available at: http://www.sciencedirect.com/science/article/pii/S0951832016306950.
5. Aven, T., 2016. On the use of conservatism in risk assessments. *Reliability Engineering and System Safety*, 146, pp.33–38. Available at: http://dx.doi.org/10.1016/j.ress.2015.10.011.
6. Aven, T., 2013b. Practical implications of the new risk perspectives. *Reliability Engineering and System Safety*, 115, pp.136–145.
7. Aven, T. & Krohn, B.S., 2014. A new perspective on how to understand, assess and manage risk and the unforeseen. *Reliability Engineering & System Safety*, 121, pp.1–10. Available at: http://www.sciencedirect.com/science/article/pii/S0951832013002159.
8. Aven, T. & Ylönen, M., 2016. Safety regulations: Implications of the new risk perspectives. *Reliability Engineering & System Safety*, 149, pp.164–171. Available at: http://www.sciencedirect.com/science/article/pii/S0951832016000168.
9. Bani-mustafa, T. et al., 2018. Strength of Knowledge Assessment for Risk Informed Decision Making. In *Esrel*. Trondheim.
10. Berner, C. & Flage, R., 2016. Strengthening quantitative risk assessments by systematic treatment of uncertain assumptions. *Reliability Engineering and System Safety*, 151, pp.46–59.
11. Bjerga, T. & Aven, T., 2015. Adaptive risk management using new risk perspectives – an example from the oil and gas industry. *Reliability Engineering & System Safety*, 134, pp.75–82. Available at: http://www.sciencedirect.com/science/article/pii/S0951832014002531.
12. Bjerga, T., Aven, T. & Zio, E., 2014. An illustration of the use of an approach for treating model uncertainties in risk assessment. *Reliability Engineering and System Safety*, 125, pp.46–53.
13. Boone, I. et al., 2010. A method to evaluate the quality of assumptions in quantitative microbial risk assessment. *Journal of Risk Research*, 13(3), pp.337–352. Available at: http://www.scopus.com/inward/record.url?eid=2-s2.0-77951165131&partnerID=40&md5=12a3caae6ff5f3fae9967becb6b35f17.
14. Borgonovo, E. & Cillo, A., 2017. Deciding with Thresholds: Importance Measures and Value of Information. *Risk Analysis*, 37(10), pp.1828–1848.
15. Cacuci, D.G., Ionescu-Bujor, M. & Navon, I.M., 2003. *Sensitivity and uncertainty analysis*, Chapman & hall/CRC Boca Raton, Florida.
16. Christopher Frey, H. & Patil, S.R., 2002. Identification and review of sensitivity analysis methods. *Risk*

analysis, 22(3), pp.553–578.

17. Commission, U.S.N.R., 1995. Use of probabilistic risk assessment methods in nuclear activities: Final policy statement. *Federal Register*, 60, p.42622.
18. Davies, M., 2015. Knowledge–Explicit, implicit and tacit: Philosophical aspects. *International encyclopedia of the social & behavioral sciences*, pp.74–90.
19. Downing, D.J., Gardner, R.H. & Hoffman, F.O., 1985. An examination of response-surface methodologies for uncertainty analysis in assessment models. *Technometrics*, 27(2), pp.151–163.
20. Duménigo, C. et al., 2008. Risk analysis methods: their importance for safety assessment of practices using radiation. In *Proceedings Congress IRPA-2008*.
21. Eiser, J. et al., 2012. Risk interpretation and action: A conceptual framework for responses to natural hazards. *International Journal of Disaster Risk Reduction*, 1(1), pp.5–16.
22. EPRI, 2015. *An Approach to Risk Aggregation for RisK-Informed Decision-Making*, Palo Alto, California.
23. EPRI, 2012. *Practical Guidance on the Use of Probabilistic Risk Assessment in Risk-Informed Applications with a Focus on the treatment of Uncertainty*, Palo Alto, California.
24. Ferdous, R. et al., 2013. Analyzing system safety and risks under uncertainty using a bow-tie diagram: An innovative approach. *Process Safety and Environmental Protection*, 91(1), pp.1–18. Available at: http://www.sciencedirect.com/science/article/pii/S0957582011000954.
25. Flage, R. & Aven, T., 2009. Expressing and communicating uncertainty in relation to quantitative risk analysis. *Reliability: Theory & Applications*, 4(2–1 (13)).
26. Goerlandt, F. & Montewka, J., 2014. Expressing and communicating uncertainty and bias in relation to Quantitative Risk Analysis. *Safety and Reliability: Methodology and Applications*, 2(13), pp.1691–1699. Available at: http://www.crcnetbase.com/doi/abs/10.1201/b17399-230.
27. Hamby, D.M., 1994. A review of techniques for parameter sensitivity analysis of environmental models. *Environmental monitoring and assessment*, 32(2), pp.135–154.
28. IAEA, 2006. *Determining the Quality of Probabilistic Safety Assessment (PSA) for Applications in Nuclear Power Plants*, Vienna: INTERNATIONAL ATOMIC ENERGY AGENCY. Available at: http://www-pub.iaea.org/books/IAEABooks/7546/Determining-the-Quality-of-Probabilistic-Safety-Assessment-PSA-for-Applications-in-Nuclear-Power-Plants.
29. INSAG, 2011. *A Framework for an Integrated Risk Informed Decision Making Process*, Vienna: INTERNATIONAL ATOMIC ENERGY AGENCY. Available at: http://www-pub.iaea.org/books/IAEABooks/8577/A-Framework-for-an-Integrated-Risk-Informed-Decision-Making-Process.
30. Karanki, D.R. et al., 2009. Uncertainty Analysis Based on Probability Bounds (P-Box) Approach in Probabilistic Safety Assessment. *Risk Analysis*, 29(5), pp.662–675. Available at: http://dx.doi.org/10.1111/j.1539-6924.2009.01221.x.
31. Khorsandi, J. & Aven, T., 2017. Incorporating assumption deviation risk in quantitative risk assessments: A semi-quantitative approach. *Reliability Engineering & System Safety*, 163, pp.22–32.
32. Kloprogge, P., Van der Sluijs, J.P. & Petersen, A.C., 2011. A method for the analysis of assumptions in model-based environmental assessments. *Environmental Modelling and Software*, 26(3), pp.289–301.

Available at: http://dx.doi.org/10.1016/j.envsoft.2009.06.009.

33. Nicolas Zweibaum & Jean-Pierre Sursock, 2014. *Addressing multi-hazards risk aggregation for nuclear power plantsthrough response surface and risk visualization*, Palo Alto, California.
34. NRC, 2011. *AN APPROACH FOR USING PROBABILISTIC RISK ASSESSMENT IN RISK-INFORMED DECISIONS ON PLANT-SPECIFIC CHANGES TO THE LICENSING BASIS*,
35. NRC, 2010. *Reactor Coolant System and Connected Systems*, Washington: NRC.
36. Oberkampf, W.L., Pilch, M. & Trucano, T.G., 2007. Predictive capability maturity model for computational modeling and simulation. *cfwebprod.sandia.gov*. Available at: https://cfwebprod.sandia.gov/cfdocs/CCIM/docs/Oberkampf-Pilch-Trucano-SAND2007-5948.pdf%5Cnfile:///Users/markchilenski/Documents/Papers/2007/cfwebprod.sandia.gov%0A/Oberkampf/cfwebprod.sandia.gov%0A 2007 Oberkampf.pdf%5Cnpapers://31a1b09a-25a9-4e20-879d-4.
37. Perhac Jr, R.M., 1996. Does risk aversion make a case for conservatism. *Risk*, 7, p.297.
38. Reinert, J.M. & Apostolakis, G.E., 2006. Including model uncertainty in risk-informed decision making. *Annals of Nuclear Energy*, 33(4), pp.354–369. Available at: http://www.sciencedirect.com/science/article/pii/S0306454905002781.
39. Riesch, H., 2013. Uncertainty. In *Essentials of Risk Theory*. pp. 29–57. Available at: http://link.springer.com/10.1007/978-94-007-5455-3.
40. Saaty, T.L., 2008. Decision making with the analytic hierarchy process. *International Journal of Services Sciences*, 1(1), p.83.
41. Simola, K. & Pulkkinen, U., 2004. *Risk Informed Decision Making A Pre-Study*, Finland: Nordisk Kernesikkerhedsforskning.
42. Siu, N. et al., 2015. FIRE PRA MATURITY AND REALISM: A DISCUSSION AND SUGGESTIONS FOR IMPROVEMENT.
43. Van Der Sluijs, J.P. et al., 2005. Combining Quantitative and Qualitative Measures of Uncertainty in Model-Based Environmental Assessment: The NUSAP System. *Risk Analysis*, 25(2), pp.481–492. Available at: http://doi.wiley.com/10.1111/j.1539-6924.2005.00604.x.
44. Talisayon, S.D., 2009. Monitoring and evaluation in knowledge management for development. *IKM Emergent Paper*, 3.
45. Veland, H. & Aven, T., 2015. Improving the risk assessments of critical operations to better reflect uncertainties and the unforeseen. *Safety Science*, 79, pp.206–212. Available at: http://www.sciencedirect.com/science/article/pii/S092575351500154X.
46. Viscusi, W.K., Hamilton, J.T. & Dockins, P.C., 1997. Conservative versus mean risk assessments: Implications for Superfund policies. *Journal of environmental economics and management*, 34(3), pp.187–206.
47. Walker, W.E. et al., 2003. Defining Uncertainty: A Conceptual Basis for Uncertainty Management in Model-Based Decision Support. *Integrated Assessment*, 4(1), pp.5–17. Available at: https://doi.org/10.1076/iaij.4.1.5.16466.
48. Whipple, C., 1987. Dealing with uncertainty about risk in risk management. In *Risk Assessment and Management*. Springer, pp. 529–536.
49. Wynne, B., 1992. Uncertainty and environmental learning: Reconceiving science and policy in the

preventive paradigm. *Global Environmental Change*, 2(2), pp.111–127. Available at: http://www.sciencedirect.com/science/article/pii/0959378092900172.

50. Zio, E., 1996. On the use of the analytic hierarchy process in the aggregation of expert judgments. *Reliability Engineering and System Safety*, 53(2), pp.127–138.
51. Zio, E. & Pedroni, N., 2012. *Overview of risk-informed decision-making processes*, FonCSI.

Appendix III (P3):

An extended method for evaluating assumptions deviations in quantitative risk assessment and application to external flooding risk assessment of a nuclear power plant

An extended method for evaluating assumptions deviations in quantitative risk assessment and application to external flooding risk assessment of a nuclear power plant

Tasneem Bani-Mustafa [(1)], R. Flage [(2)], Dominique Vasseur [(3)], Zhiguo Zeng [(1)] and Enrico Zio [(1) (4)]

[(1)] *Chair on System Science and the Energetic Challenge, EDF Foundation*

Laboratoire Genie Industriel, CentraleSupélec, Université Paris-Saclay,

3 Rue Joliot Curie, 91190 Gif-sur-Yvette, France

[(2)] *University of Stavanger, Norway*

[(3)] *EDF R&D, PERICLES (Performance et prévention des Risques Industriels du parc par la simulation et les Etudes) EDF Lab Paris Saclay - 7 Bd Gaspard Monge, 91120 Palaiseau, France*

[(4)] *Energy Department, Politecnico di Milano, Via Giuseppe La Masa 34, Milan, 20156, Italy*

Abstract

In quantitative risk assessment, assumptions are typically made, based on best judgement, conservative, or (sometimes) optimistic judgments. Best judgment and optimistic assumptions may result in failing to meet the quantitative safety objectives, whereas conservative assumptions may increase the margins which the objectives are met with but result in cost-ineffective design or operation. In the present paper, we develop an extended framework for evaluating the criticality (risk) that deviations from the assumptions made in the risk assessment lead to a reduction of the safety margins. The framework is, then applied within the quantitative risk assessment of a Nuclear Power Plant (NPP) exposed to external flooding. Compared to previous works on the subject, we consider also conservative assumptions and introduce decision flow diagrams to support the classification of the criticality of the assumptions made. We find that the framework provides a solid decision basis and that the decision flow diagrams facilitate the standardization of the evaluation of the assumption deviation effects on risk assessment.

Keywords

Quantitative risk assessment; conservative assumption; assumption deviation; strength of knowledge; decision flow diagram; nuclear power plants; external flooding.

1. Introduction

Making assumptions is an inevitable part of quantitative risk assessment (QRA) process. An assumption can be defined generally as "a fact or statement (such as a proposition, axiom […], postulate, or notion) taken for granted" (Merriam-Webster). Other definitions, from the scientific literature and more specific to the risk assessment context, include "conditions/inputs that are fixed in the assessment but which are acknowledged or known to possibly deviate to a greater or lesser extent in reality" (Berner & Flage, 2016 p. 46) and the following, which relies on the definition of defaults (Suter *et al*., 2007 pp. 134-135):

"Defaults are functional forms or numerical values that are assigned to certain models or parameters in risk assessment, based on guidance and standard practice, in the absence of good data. […] Assumptions are equivalent to defaults but are derived for a specific assessment rather than being taken from guidance. They may be complex, implying functional forms or sets of parameters. […] Ad hoc assumptions must be individually justified."

The latter definition restricts assumptions to having a quantitative format, whereas the former definitions allow also for qualitative assumptions and highlight the potential, or even expected, non-true nature of assumptions. Some examples of types of assumptions in risk assessment are:

1. The number of people exposed to a hazard
2. The reliability of a safety barrier
3. The behavior of people leading up to or following an accidental event.

The first two types of assumptions concern directly risk model parameters. If N and p denote the number of people exposed to the hazard and the reliability of the safety barrier, respectively, then the assumptions specify the numerical values of N and p. If time-dependent, the assumptions specify functional forms $N(t)$ and $p(t)$ for a time index *t*. The last assumption is likely to be more qualitative in nature, e.g. assuming that all people involved in the accidental event follow the emergency preparedness plan. Transforming this qualitatively formulated assumption into a quantitative format is less straightforward.

Risk assessment assumptions are typically of best judgement or conservative. Best judgement assumptions are here understood as reflecting the best knowledge on the matter, e.g. a realistic "best estimate" of a risk model parameter, whereas conservative assumptions come from lack of knowledge on the matter or conscious simplification of its analysis, and define conditions or values that are in some sense 'unfavorable', 'protective, with respect to the current knowledge and lack of. Optimistic assumptions are also possible, but are typically rare in risk assessment, from the safety perspectives. With reference to the above three example assumptions, a best judgement assumption would amount to considering that the number of people exposed to a hazard at a given workplace is equal to the number of employees: the actual number could deviate, e.g. due to the absence from work by some employees or due to the presence of visitors, but nonetheless, if forced to specify a single value, the number of employees would be perhaps the best justifiable choice. A conservative assumption would be that a specific safety barrier will always fail, i.e., reliability equal to zero, $p = 0$. An optimistic assumption would be that in case of an accident, all

people involved would behave according to some emergency preparedness plan.

For best judgement and optimistic assumptions, deviations of the actual conditions could cause the safety objectives to be actually unmet. With regards to this, the concept of assumption deviation risk assessment was coined by Aven (2013) to address this type of "risk" situation to evaluate different intensities of deviations, their associated probabilities of occurrence, the effect of the deviations on the consequences and an overall strength of knowledge judgement for these three attributes. Assumption deviation risk assessment, thus, goes beyond sensitivity analysis, which tends to be focused on "what if" questions, as discussed by Khorsandi & Aven (2017). In the case of conservative assumptions, on the other hand, deviations might decrease the margins for meeting the objectives.

In the present paper, we take the recently suggested method for evaluating the risk from assumptions deviations by (Khorsandi & Aven, 2017) and apply it to the external flooding risk assessment of a nuclear power plant (NPP). In doing this, we extend the overall methodology to evaluate also the risk of deviations from conservative assumptions and introduce decision flow diagrams for the quantitative evaluation. We find that the proposed extensions provide a more solid decision making basis than focusing only on best judgement assumptions and that the decision flow diagrams facilitate a standardization of the evaluation of the risk from assumptions deviations.

Works closely related to the present paper include the already mentioned papers by Aven (2013), introducing the concept of assumption deviation risk, and by Khorsandi & Aven (2017), presenting how to integrate an assumption deviation risk assessment as part of a quantitative risk assessment (QRA). Berner & Flage (2016) also build on the assumption deviation risk concept and develop a framework comprising six classes of uncertain assumptions, which is used to prescribe strategies for treating these assumptions both in the risk assessment (Berner & Flage, 2016) and in the subsequent risk management (Berner & Flage, 2017).

The remainder of the paper is organized as follows. In Section **Error! Reference source not found.**, we describe the extended method. Then, in Section **Error! Reference source not found.**, we present the application to the case study. In Section **Error! Reference source not found.**, we offer a discussion of some conclusions.

2. Extended framework for the evaluation of assumptions deviations

In this section, we extend the original work of Khorsandi and Aven (2017) for a more comprehensive assessment of the criticality (risk) of assumptions deviations. In Sect. 2.1, we present the extended framework and compare it to the original one. In Sect. 2.2, the detailed implementation of the framework is described.

2.1. The assessment framework

In this section, the original work of Khorsandi and Aven (2017) is extended considering multiple contexts of decision-making and multiple types of assumptions. We assume that each assumption As_i affects the numerical values of some parameters in the Probabilistic Risk Assessment (PRA) model. The factor that links the assumptions to the numerical parameters is called "juncture" in this paper. The

criticality (C) of an assumption deviation quantifies its risk impact in terms of the likelihood of the deviation, the severity of its influence on the decision making considered and the strength of the knowledge supporting the assumption. Three levels of criticality are defined with their corresponding settings:

1. Very critical $(C = 1)$: The assumption is made based on weak knowledge and the confidence on the assigned value of the model parameters is low. The deviation is very likely to happen. Besides, the assumption deviation has severe influence on the decision making and might lead to exceedance of the safety limit. Further analysis and justification of the assumption is required.

2. Not very critical $(C = 2)$: The assumption is made based on a moderate level of knowledge. The assumption deviation is likely to happen, but the risk metric remains within the safety limits even after considering such assumption deviation. The assumption can be trusted to support decision making if the risks of the deviation from other assumptions are all not critical $(C = 3)$. Further analysis and justification of the assumption is needed only when multiple other assumptions are also in this state.

3. Not critical $(C = 3)$: The assumption made is based on strong knowledge. An assumption deviation is unlikely to happen or, if it happens, it does not affect the decision making. The assumption can be trusted and decisions can be made based on the current assumption.

To evaluate the criticality of the assumptions deviations, six criteria are considered, as shown in Figure 1:

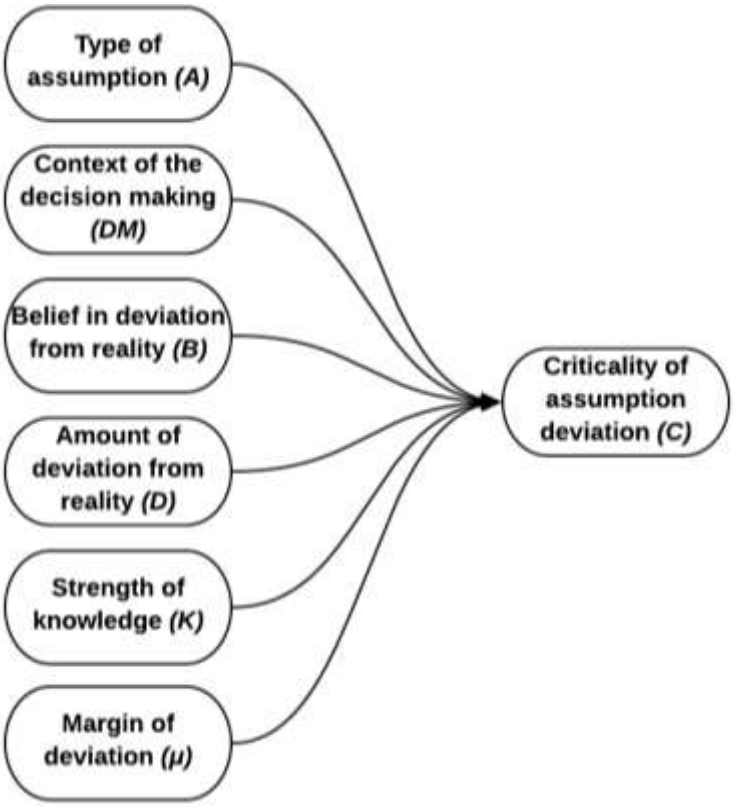


Figure 1 Criteria for evaluating the criticality of assumption deviation.

1. Type of assumption (A): Assumptions made in PRA can be classified into different types. For example, EPRI (2015) distinguishes three types of assumptions: conservative assumptions, best judgment assumptions and approximations. Conservative assumptions are made out of cautiousness and tend to overestimate the risk rather than underestimate it; best judgment assumptions are believed to represent expected scenarios, given the available knowledge; approximations are assumptions that are made for reducing the complexity of the models (EPRI 2006). Deviations in different types of assumptions might lead to different influences on the PRA. In our framework, three types of assumptions are considered:
   i. Optimistic assumption (A_1): the assumption is judged by peers to underestimate the risk when compared to reality
   ii. Best judgment (A_2): the assumption is judged by peers as representative of reality (realistic)
   iii. Conservative assumption (A_3): the assumption is judged by peers to overestimate the risk when compared to reality (pessimistic).
2. Context of the decision making (DM): Risk metrics are used to support decision making in different contexts (EPRI 2015). In this paper, we distinguish between two contexts of decision making: comparison to safety objectives, where by the risk metrics are compared to quantitative safety goals and criteria (EPRI 2015), and comparison of alternatives, where by the risk metrics of different alternatives are compared in order to make a choice among the alternatives. The criticality of assumptions deviations varies from one context to another, where, in comparing risk metric to a safety goal, only the deviation toward critical scenarios need to be considered. On the other hand, for comparing alternatives in terms of their risks, all the deviation scenarios need to be considered, since a conservative assumption might lead to a higher risk metric and hence, lead the decision maker to make a wrong decision by choosing another alternative that has a higher risk in reality but appears lower due to the different levels of conservatism in the analysis.

3. Belief in deviation (B) measures the realism of an assumption and is expressed by the likelihood of assumption deviation. The likelihood is assigned by the experts following the criteria defined in Khorsandi and Aven (2017), i.e., what could cause the assumption to deviate in reality; what are the key drivers of those causes; etc.

4. Amount of deviation from reality (D) refers to the amount of deviation between the assumed parameter value and the true value. It is assigned by experts and expressed in percentage.

5. Strength of knowledge (K) refers to the strength of the background knowledge that supports the evaluation of the belief in deviation and the amount of deviation.

6. Margin of deviation (μ) refers to the degree to which an assumption may deviate before the deviation changes the decisions made based on the results of risk assessment, e.g., the violation of the acceptance criteria or the change of the prioritization of different options. This margin is calculated analytically (see Sect. 2.2.8) and expressed in percentage.

The logical combination of the six criteria yields different levels of criticality. Decision flow diagrams are introduced in this paper to capture the logical relationship between the six criteria and the criticality of assumptions deviations (see Sect. 2.2.9). A comparison between the original assessment framework in Khorsandi and Aven (2017) and the extended framework is made in Figure 2. It can be seen that the original work of Khorsandi and Aven (2017) is adjusted and extended to include an additional context of decision making (comparing alternatives) and also a new type of assumption (conservative assumptions). Accordingly, new criteria are added or adjusted to integrate the new decision context and type of assumption in the assessment of the assumption deviation risk. As to the presentation of the assumption deviation risk, the radar plot in Khorsandi and Aven (2017), which presents the contributing factors to the assumption deviation risk individually, is replaced with an overall integrated metric for assumption deviation risk, i.e., the criticality (C). These extensions make it possible for the extended framework to provide a more comprehensive description of the risk from assumptions deviations.

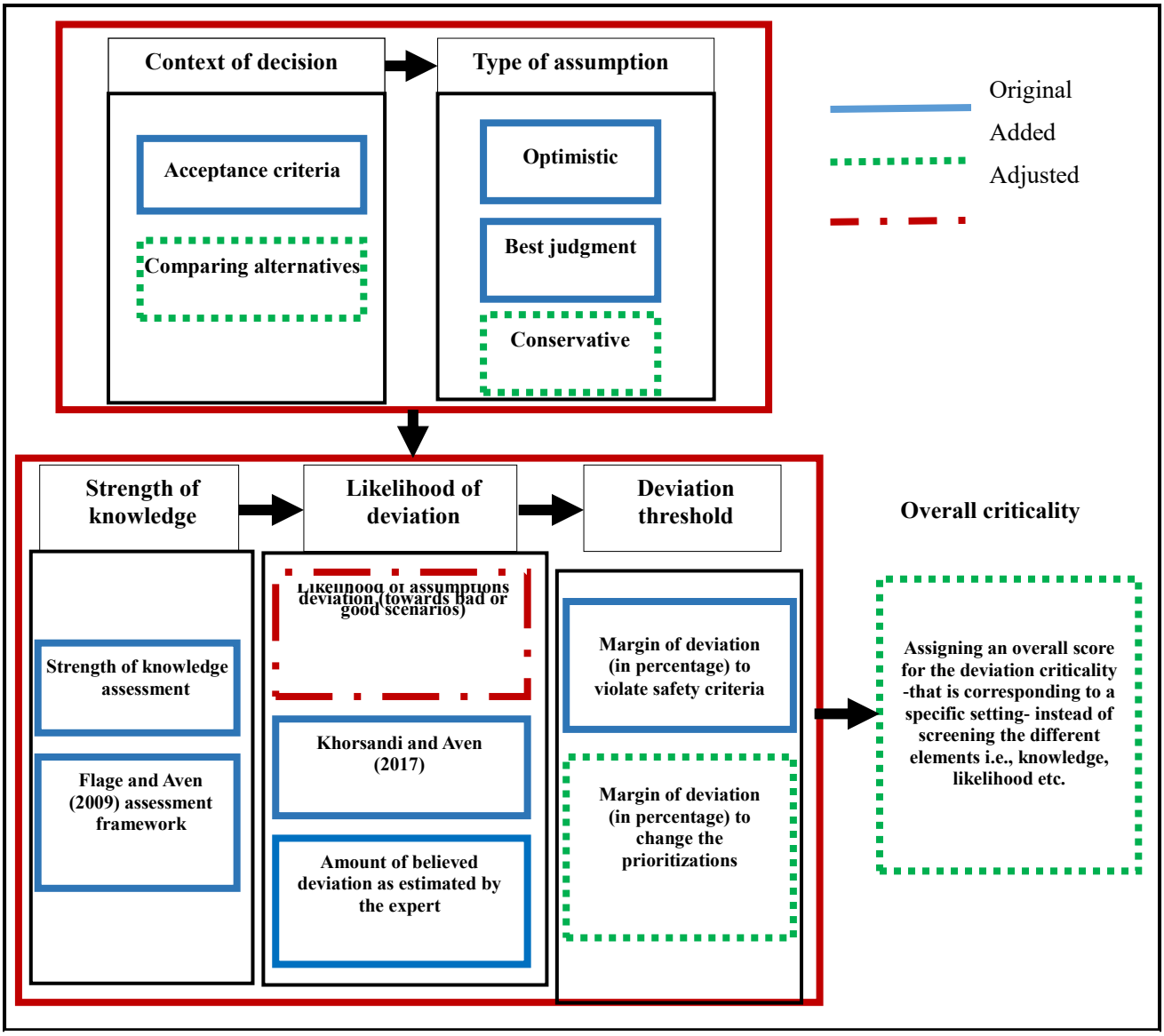


Figure 2 A comparison between the original (Khorsandi & Aven, 2017) and the extended frameworks for assumption deviation risk assessment.

2.2. Implementation of the framework

As shown in Figure 3, nine main steps are needed for applying the developed framework to assess the criticality of assumptions deviations. The nine steps are discussed in details in sub Sect. 2.2.1-2.2.9.

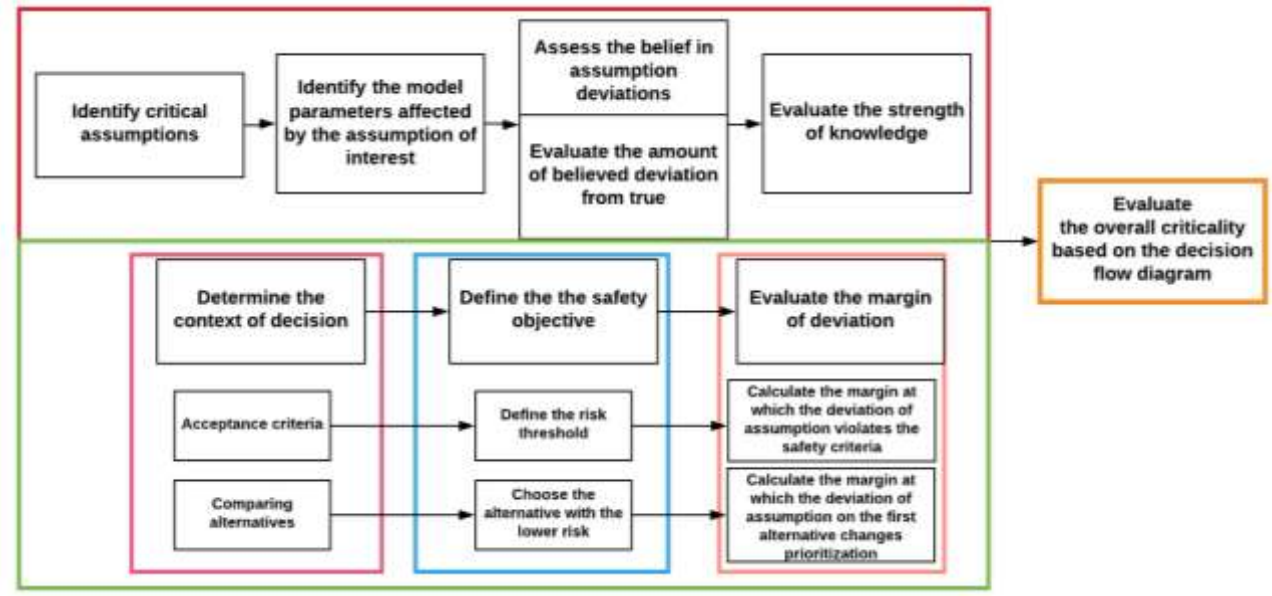


Figure 3 Procedure for applying the developed framework for assumption deviation criticality (risk) assessment.

2.2.1. Identify critical assumptions

In the first step, the assumptions made in the PRA are identified. The assumptions might be made due to lack of understanding and knowledge about a phenomenon or as an attempt to reduce the modeling details and complexity (EPRI 2006), (EPRI 2015). The type of each assumption (A) is determined by expert judgment, making reference to the definitions in Sect 2.1.

2.2.2. Identify the model parameters affected by the assumption of interest

As mentioned in Sect 2.1, in this paper, we assume that there is a juncture that connects numerically an assumption to one or more parameters in the PRA model. Without losing generality, let us assume that the PRA model is represented by:

$$R = f(p_1, p_2, \dots p_m, \dots p_n), \tag{1}$$

where R is the risk metric and $p_1, p_2, \dots p_m, \dots p_n$ are the model parameters (e.g., failure probabilities). The juncture can be conceptually represented as Figure 11, where As represents a set of assumptions. The second step, then, involves identifying the model parameters affected by each assumption, as shown in Figure 11.

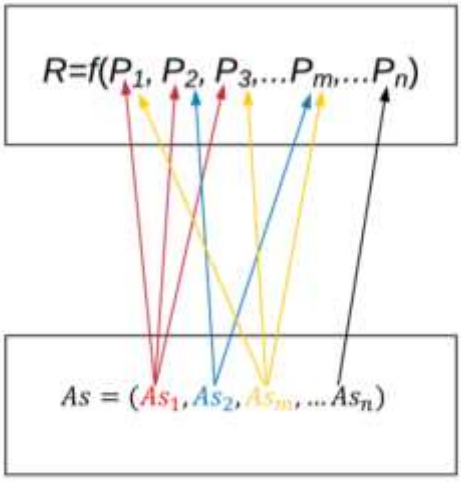


Figure 4 Representation of connections between assumptions and model parameters.

2.2.3. Assess the belief in assumption deviation

The belief in deviation is evaluated as the subjective probability assigned by experts that the assumption deviates from the actual conditions. The assigned value is conditional on the available background knowledge, including experts' individual expertise. It should be noted that the aim of evaluating the belief in deviation is not to assign a precise value for the probability of deviation. Rather, it aims at expressing the experts' beliefs , based on the available knowledge, on how likely the assumption might be deviating from reality (Khorsandi and Aven 2017). Such a step can be regarded as a tool for making good use of experts' individual expertise by reflecting their implicit knowledge that cannot be directly stated or documented.

To determine the value of B, the likelihood (l) needs to be evaluated by experts first, following the considerations recommended by Khorsandi and Aven (2017): What could cause the assumption to deviate? What are the key drivers of those causes? Has a similar deviation occurred in the past? What evidence is available for supporting the potential for a deviation?

Then, the value of B is determined based on the likelihood (l):

a. $B = 1, if\ 0 \leq l \leq 20\%$
b. $B = 2, if\ 20\% < l \leq 30\%$
c. $B = 3, if\ 30\% < l \leq 100\%$

2.2.4. Evaluate the amount of believed deviation from the true value

The amount of believed deviation is evaluated as the relative distance between the assumed parameter value and the true value believed by experts, as expressed by Eq. (2). Similar to the belief in deviation, the believed deviation D is evaluated by experts and represents the experts' belief on how severe the deviation could be. The value assigned to D takes a positive sign $(+)$ if the assumption is believed to deviate towards dangerous scenarios and a negative sign $(-)$ if it is deviating towards safe scenarios:

$$D = \frac{p_t - p}{p} \tag{2}$$

where D is the amount of believed deviation, p_t is the parameter value believed true by the experts, and p is the parameter value as assumed in the analysis.

2.2.5. Evaluate the strength of knowledge

The assigned belief (likelihood) and amount of deviation are conditional on the background knowledge available, and on the individual expertise and points of view of the experts who made the assessment. Therefore, the strength of knowledge on which the assessment is based is highly relevant and is explicitly considered in both the original and extended framework. In this paper, we use the method proposed by Flage and Aven (2009) for evaluating the strength of knowledge. This approach is mainly based on the evaluation of four criteria: (i) reasonability and realism of assumptions; (ii) phenomenological understanding; (iii) availability of reliable data and information; (iv) agreements among peers. In addition, we take into account a fifth criteria, suggested by Khorsandi and Aven (2017): (v) the level of expertise and competence of the experts. A score of 1-3 is given for each criterion, corresponding to three levels, i.e., weak, moderate and strong, respectively.

A weighted average of the five criteria scores $k_i, i = 1,2,\dots 5,$ is used to calculate the overall knowledge score SK:

$$SK = \sum_{i=1}^{5} w_i \cdot k_i, \tag{3}$$

where w_i is the weight of criterion k_i. Obviously, the five criteria are not equally important in defining the strength of knowledge. To handle this, the Analytical Hierarchy Process (AHP) (Saaty 2008) is used to determine the weights of the strength of knowledge criteria. A good feature of the method is that it can be helpful in group decision-making (Saaty 2008). Experts are asked to fill pairwise comparison matrixes that represent the relative importances of the five criteria in defining the knowledge. The eigenvector problem is, then, solved and the weights are found by normalizing the principal eigenvector. The calculated weights from the experts are, then, averaged to give the final weights shown in Table 1.

Table 1 Strength of knowledge criteria and their weights.

Attribute	Weight
Reasonability and realism of assumptions (k_1)	0.13
Availability of reliable data and information (k_2)	0.13
Phenomenological understanding (k_3)	0.42
Agreement among peers (k_4)	0.16
Level of expertise and competence of the experts (k_5)	0.16

The strength of knowledge denoted by K, is, then, calculated based on the value of SK:

- $K = 1,$ if $1 \leq SK \leq 1.6$
- $K = 2,$ if $1.6 < SK \leq 2.3$
- $K = 3,$ if $SK > 2.3$

2.2.6. Determine the context of decision

In the original work of Khorsandi and Aven (2017), only one context of decision making was considered, i.e., comparing a risk metric to a specific safety objective. In this sense, only assumptions deviations toward dangerous scenarios need to be considered. In the practice of risk management, however, we often need to compare alternatives in terms of their risks. In this case, all the deviation scenarios need

to be considered, since a conservative assumption might lead to a higher risk metric, which again leads the decision maker to prefer other alternatives; in other words, it gives a "false alarm" of high risk. For more illustration, take the example in Figure 5. In this example, the decision maker is comparing two alternatives, A_1 and A_2, and he/she prefers to choose the alternative with the lower risk. At a first glance, the decision maker would choose A_1 as it has the lowest risk metric value (the blue solid line). However, a second look shows that the value of R_2 (in the meshed filling) is lower than that of R_1, when the true condition is used in the calculation rather than a conservative assumption.

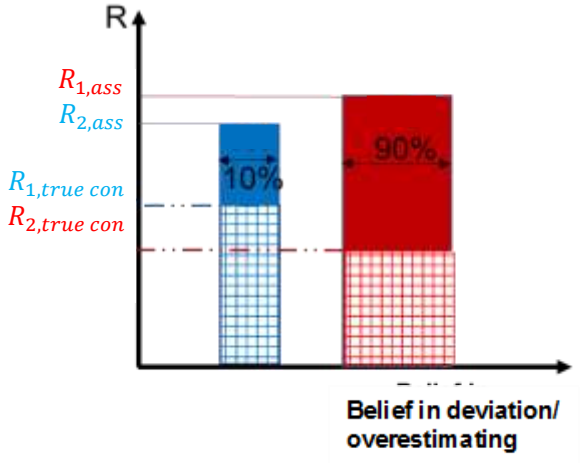


Figure 5 Comparing the risk related to two alternatives taking into account the risk metric value based on the assumption made and the true condition.

Hence, it is important to identify the context of decision making when implementing the extended framework. In this paper, two decision making contexts are distinguished, namely, comparing a risk metric to a safety objective (DM_1) and comparing two alternatives (DM_2).

2.2.7. Define the safety objective

The safety objective needs to be identified considering the given decision context, as shown in Figure 3. The safety objective represents a numerical value whose exceedance by the risk metric would lead to changes in the results of the risk-informed decision making. The safety objective is dependent on the context of the decision making. For the decision context DM_1, the safety objective is identified as the threshold that the risk metric should not exceed. On the other hand, if the decision context is DM_2, the assessor needs to choose the alternative with the lowest risk metric value. Therefore, the (higher) risk metric value of another alternative is defined as the safety objective under this decision making context.

2.2.8. Identify the margin of deviation

Next, the margin of deviation (μ) needs to be calculated. This margin represents the maximum tolerable assumption deviation before the risk-informed decision is changed. As shown in Figure 11, different assumptions might affect one or more model parameters, or, the other way around, a model parameter might be affected by one or more assumptions. In this paper, we calculate the margin of deviation one assumption at a time, to reduce the complexity of the analysis. Assume that the assumption of interest a_i affects model parameters $p_1,\ p_2, \ldots p_m$. Then, we assume that the influence of the assumption

deviation on the $p_1,\ p_2, \dots p_m$ can be modeled by:

$$\begin{cases} \acute{p_1} = (1+\mu)p_1 \\ \acute{p_2} = (1+\mu)p_2 \\ \vdots \\ \acute{p_m} = (1+\mu)p_m \end{cases} \tag{4}$$

where $\acute{p_i},\ \ i = 1, 2, \dots m,$ are the deviated model parameters and μ represents the amount of deviation in the model parameters due to the deviation in the assumption. Then, the deviated risk metric $\acute{R}$ is calculated by

$$\acute{R} = f(\acute{p_1}, \acute{p_2}, \dots \acute{p_m}, p_{m+1} \dots p_n) \tag{5}$$

The value of μ can be calculated by solving the following equation:

$$\underset{\mu}{arg}\ f\left((1+\mu)\cdot p_1, (1+\mu)\cdot p_2, \dots (1+\mu)\cdot p_m,\ p_{m+1}, \dots p_n\right) = R_{th} \tag{6}$$

In Eq. (6), R_{th} is the safety objective defined in Sect. 2.2.7, i.e.:

$$R_{th} = \begin{cases} R_{lim}, if\ the\ decsion\ context\ is\ DM_1 \\ R_2, if\ the\ decsion\ context\ is\ DM_2 \end{cases} \tag{7}$$

where R_{lim} and R_2 represent the safety limit objective and the risk metric value of the alternative being compared, respectively.

2.2.9. Evaluate the overall criticality based on the decision flow diagrams

The criticality of an assumption deviation measures its influence on the risk-informed decision making and, hence, on the safety of the system. As defined in Sect. 2.1, the criticality of the assumption deviation depends on both the severity of the influence and the likelihood of the deviation. Four scenarios are distinguished to quantify the severity of the influence of the assumption deviation:

a. failures in meeting the established objectives, i.e., the magnitude of deviation is larger the deviation margin, leading to the exceedance of the safety limit;
b. success in meeting the established objectives i.e., the magnitude of deviation is lower than the deviation margin, or the deviation is occurring towards lower amounts of risk due to conservatism in the assumption;
c. Altering the different prioritization when comparing two or more alternatives, i.e., the risk metric based on unrealistic assumptions is higher or lower than what it would be based on the true conditions, leading to the mischoice among the different alternatives.
d. Unchanging the prioritization when comparing two or more alternatives, i.e., the risk metric based on unrealistic assumptions is higher or lower than what it would be based on the true conditions, leading to misranking the different alternatives.

Considering the scenarios defined above and the likelihood of deviation, decision flow diagrams are built in Figure 6-8 for evaluating the criticality of assumption deviation risk. It should be noted that in these figures, the difference between the margin of deviation μ and the amount of deviation D, denoted by $\Delta\mu$, is calculated and used to measure the safety margin for a given assumption deviation:

$$\Delta\mu = \mu - D \tag{8}$$

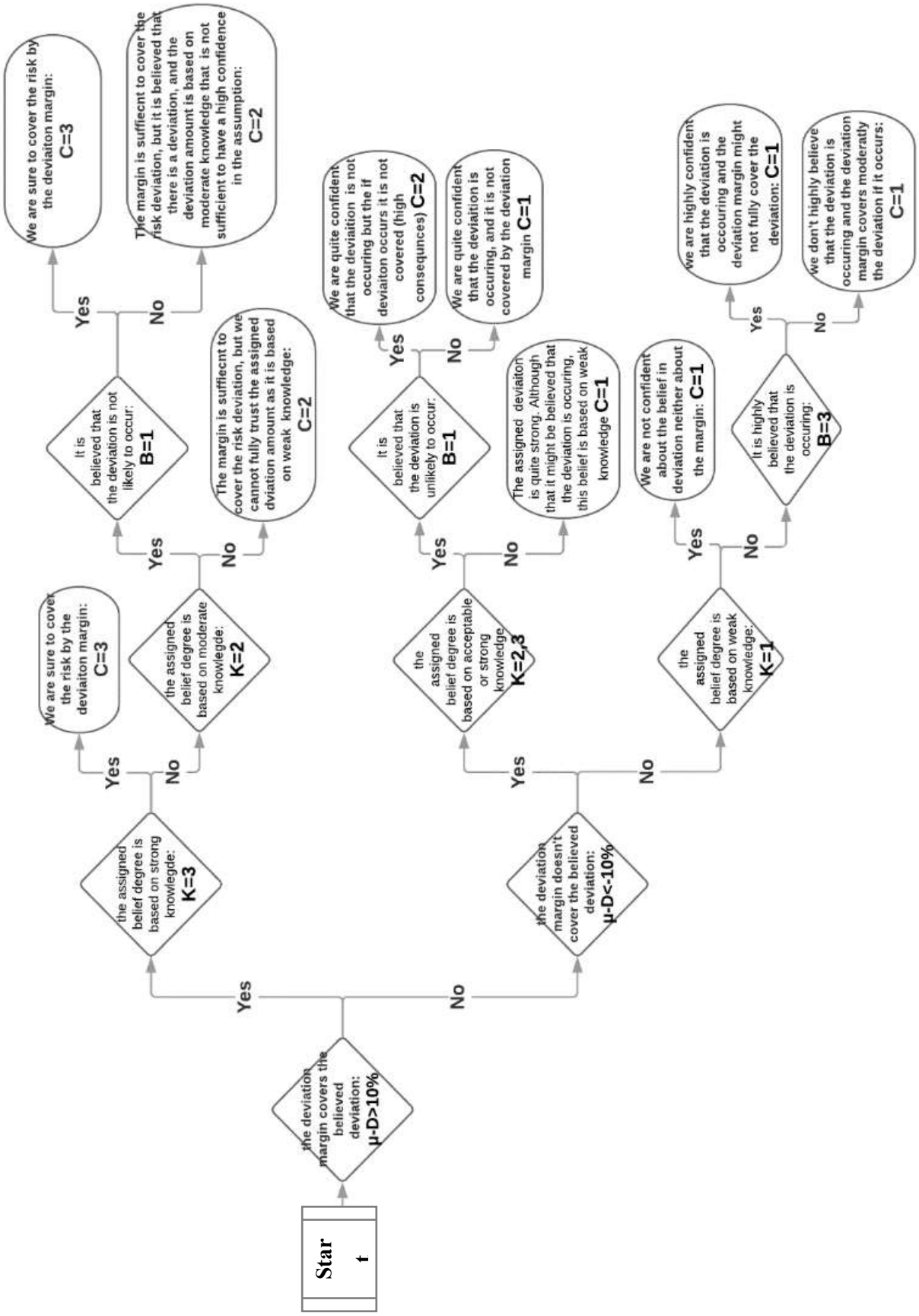


Figure 6 Criticality assessment decision flow diagram for decision context DM_1 and assumptions of types A_1 and A_2.

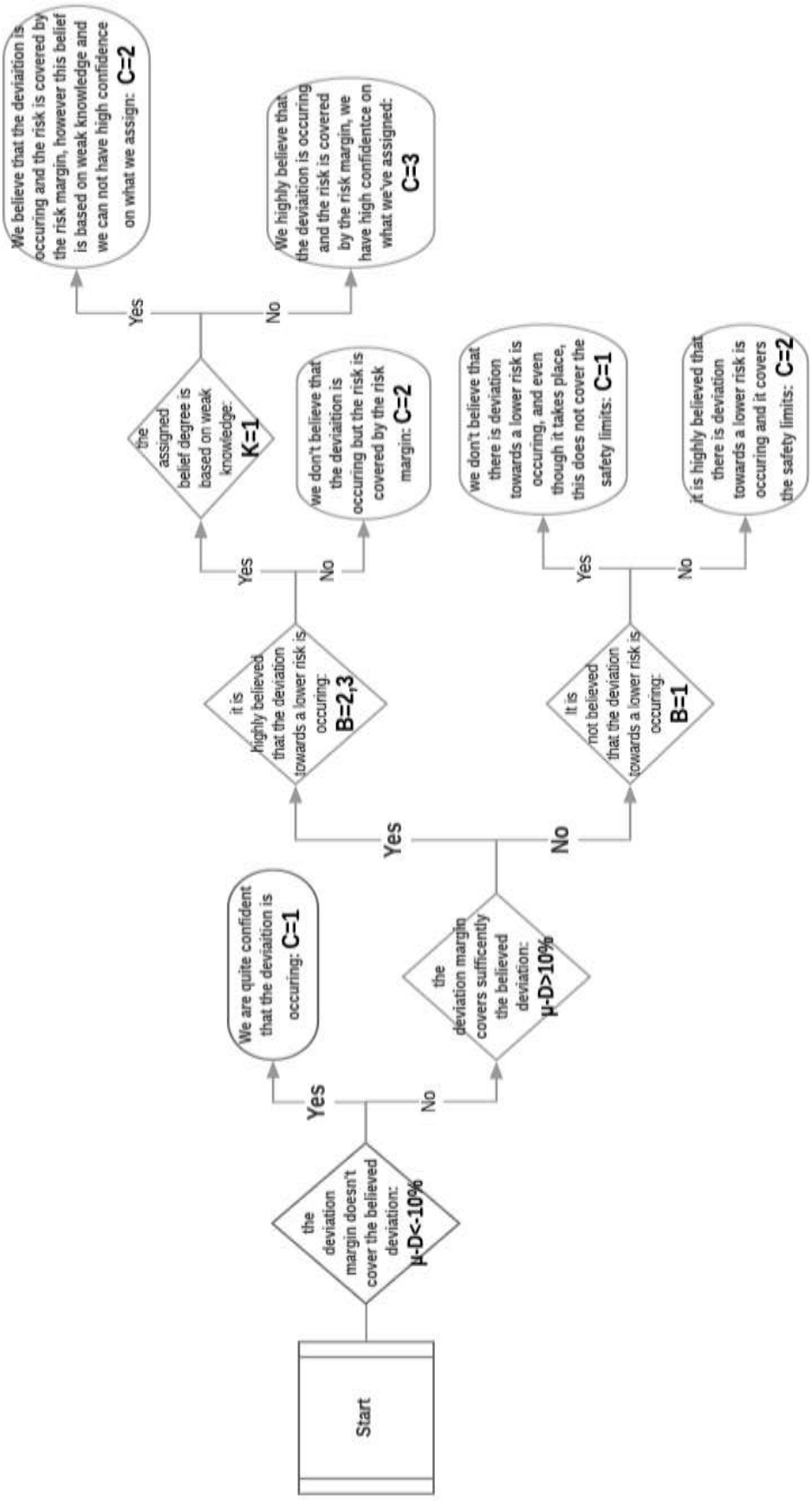


Figure 7 Criticality assessment decision flow diagram for decision context DM_1 and assumptions of type A_3.

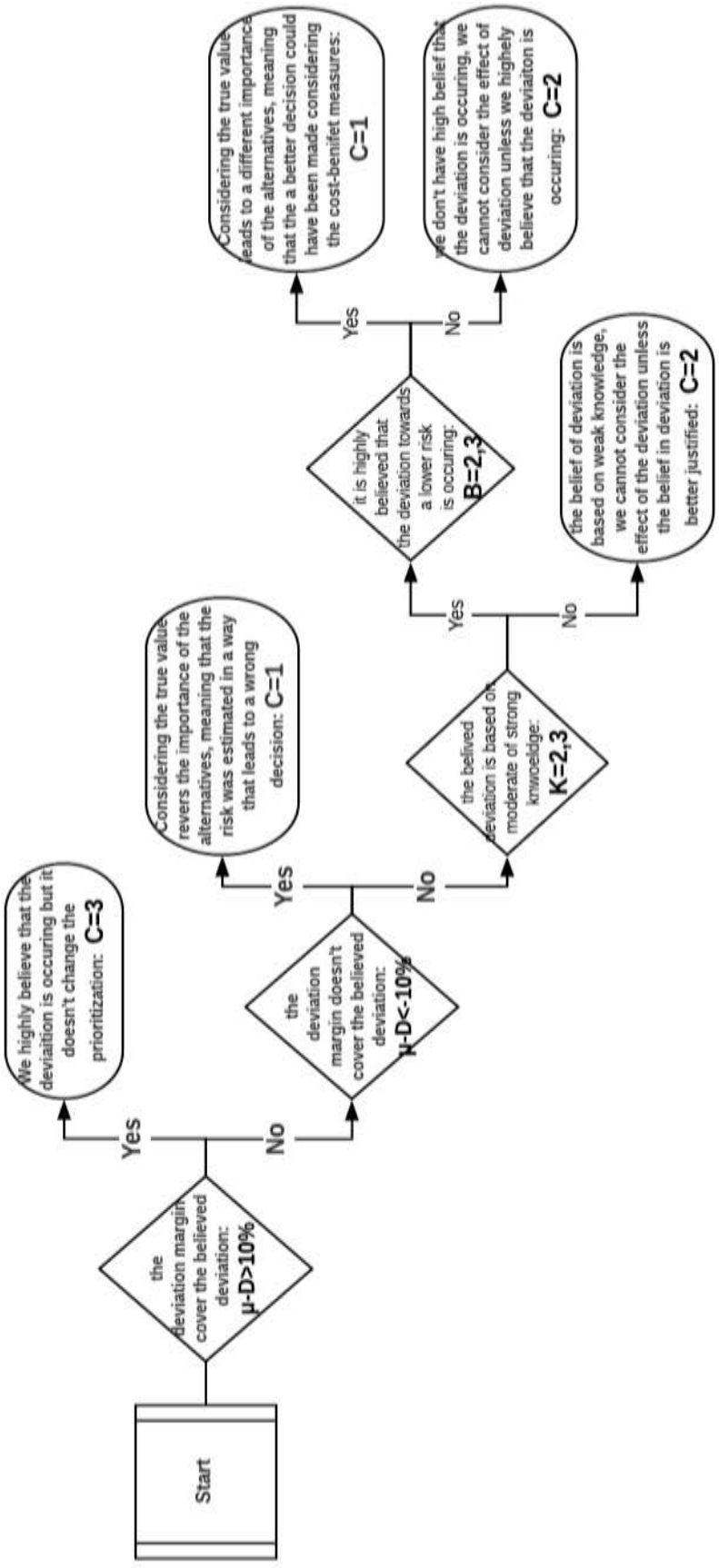


Figure 8 Criticality assessment decision flow diagram for decision context DM_2 and assumptions of types A_1, A_2 and A_3.

Following the steps in Sects. 2.2.1-2.2.7, the criticality C can be evaluated using the decision flow

diagrams in Figures 6-8. Take the case in Table 2 as an illustrative example. In this example, the assessor assigns a 90% probability of deviation, meaning that he or she is almost sure that the assumption deviates from reality. The amount of the believed deviation is evaluated to be 20%. The two values are assigned based on strong knowledge, i.e., $k = 3$, which means that the assessment is judged to be credible to a certain degree and can be trusted. The difference between the deviation margin and the amount of the believed deviation is 40%. This logically means that we are sure to be under the safety limits even though the real condition deviates from the assumption. However, as the decision context in this example is DM_1 and the type of assumption is A_2, the decision flow diagram in Figure 6 is chosen for evaluating C. It can be seen from Figure 6 that in this case, we have $C = 3$, meaning that the assumption can be trusted and that decisions can be made based on the current assumption, as the assumption deviation risk is judged to be low.

Table 2 An example of a classification of assumptions deviation risk.

Criteria	Assessment
Type of assumption (A_i)	Best judgment
Context of decision making (DM_i)	Comparing the risk metric to a risk limit
Likelihood of deviation (l)	90%
Amount of believed deviation (D)	20%
Strength of knowledge (K)	Strong
Margin of deviation (μ)	60%

3. Case study

In this section, we apply the developed framework on a case study of real PRA models for the external flooding hazard groups in an NPP. The PRA models were developed by Electricité de France (EDF). The needed data and information that supports the model development were found in the technical reports provided by EDF, which are not mentioned here for confidentiality reasons.

3.1. Description of the PRA model

PRA models are used for investigating undesired events and quantifying their likelihoods and consequences. Similar to all analytical models, PRA models are conditional on the models' assumptions (EPRI 2015). The assumption made are mainly: (i) assumptions made in case of lack of information and understanding of some phenomena or risk-related aspects; (ii) assumptions made for reducing the complexity of the model and to make it operational (these assumptions are also called approximations in (EPRI 2015)). The PRA model for external flooding is chosen because it is less mature compared to the PRA model of other hazard groups and involves many assumptions.

External flooding is a naturally induced hazard that might be caused due to different initiating events, such as river overflows, dam failures and snow melts (IAEA 2003), (IAEA 2011). The PRA model developed by EDF is a combination of fault trees and event trees, evaluated under different scenarios, e.g., water levels and operation states. The model structure and the probabilities of basic events (BEs) are, in turn, related to specific assumptions made by experts. The original external flooding PRA model is of a large scale (i.e., it includes three operation states, thousands of BEs and several thousand Minimal Cut Sets (MCS), and a large number of assumptions). A reduced-order model has been constructed in Bani-Mustafa

et al. (2018) to represent the original model with less complexity, i.e., less BEs and less MCSs. In this paper, we consider the reduced-order model in Bani-Mustafa *et al.* (2018) for assumption deviation risk assessment. In this reduced order model, only one operating state (Normal Shutdown with cooling using Steam Generator-NS/SG) that contributes to 86% of the risk metric value is considered. In this operating state, one scenario (water levels) whose risk contribution is 98.7% is considered. Given the operating state and scenario considered, 5 MCSs that contribute to 80.1% of the risk are considered. The corresponding MCSs and BEs of the reduced-order model are presented in Tables 3-4.

Table 3 Reduced-order model constituents (Bani-Mustafa *et al.* 2018).

Operating state	Scenarios	MCS
NS/SG	Water level A	MCS1={BE1, BE2, BE3} MCS2={BE2, BE3, BE4} MCS3={BE3, BE5, BE6, BE7, BE8} MCS4={BE2, BE3, BE7, BE9} MCS5={ BE2, BE3, BE6, BE10}

Table 4 Basic events included in the reduced-order model (Bani-Mustafa *et al.* 2018).

Symbol	Basic event
BE1	External flooding with water level A inducing a loss of offsite power
BE2	Loss of auxiliary feedwater system due to the failure to close the isolating valve
BE3	Loss of component cooling system because of clogging
BE4	Failure of all pumps of the Auxiliary feedwater (AFW) system
BE5	Failure of the turbine of AFW system
BE6	Failure of the Diesel Generator A
BE7	Failure of the Diesel Generator B
BE8	Failure of the common diesel generator
BE9	Failure of pumps 1 and 2 of AFW system
BE10	Failure of pumps 2 and 3 of AFW system

Taking the rare-event approximation, the total risk metric R_{Red} of the reduced-order PRA model can be calculated by:

$$R_{Red} = \sum_{i=1}^{n_{O,Red}} \sum_{j=1}^{n_{S,Red,i}} \sum_{k=1}^{n_{MCS,Red,i,j}} \prod_{q \in MCS_{i,j,k}} P_{BE,q} \tag{9}$$

where $n_{O,Red}$ is the number of operation states in the reduced order model, $n_{S,Red,i}$ is the number of scenarios in the reduced-order model, $n_{MCS,Red,i,j}$ is the number of minimal cutsets in the reduced-order model, $P_{BE,q}$ are the probabilities of the basic events in the reduced-order model. As shown in Bani-Mustafa *et al.* (2018), using the reduced-order model allows reproducing approximately 68% of the total risk contribution.

3.2. Evaluation of assumption deviation risk

3.2.1. Identifying critical assumptions

The critical assumptions in the PRA model of external flooding are identified following the procedures in Sect. 2.2 and listed in Table 5. The assumption deviation risks for the assumptions in Table 5 need to be evaluated using the developed method in Sect 2. In the following, we illustrate in detail how to apply the developed framework on one conservative assumption, namely “the clogging accompanying

some floods is unpredictable and unfilterable". For the other assumptions, we directly give the classification results in Sect. 3.2.8.

Table 5 List of the assumptions related to the reduced-order model of the external flooding hazard group.

As_i	Description	Type	Affected basic event
As_1	It is assumed that failure to close the isolating valves for volumetric protection sealing-water proofing causes the total loss of EFWS	Conservative	BE2
As_2	If the floods occur, the clogging is certain ($P = 1$)	Best judgment	BE3
As_3	If the river flooding is accompanied with clogging, then, it is unpredictable and unfilterable	Conservative	BE3, BE4
As_4	Clogging leads to failure of Essential Services Water System (RRI component cooling system)	Best judgment	BE3, BE4
As_5	It is assumed that probabilities of a given level of flood can be calculated by extrapolating the distributions based on observed data to the extreme water flowrate (i.e., flowrates that have never occurred) and that the probabilities of floods can be taken as mean values	Best judgment	BE1
As_6	It is assumed that once the water reaches the bottom of an equipment, the equipment fails	Conservative	BE2-BE10
As_7	It is assumed that once the water level exceeds the height of the barriers, the water will enter and fill the building	Best judgment	BE2-BE10
As_8	It is assumed that unit 1 cannot get help from unit 2 and vice versa, or from the safeguard system shared between the two units	Conservative	BE8
As_9	It is assumed that the river flood can be predicted using statistical models	Optimistic	BE1
As_{10}	It assumed that once the river flood is predicted, the probability of failing to transit into the state of "repli: under control" (i.e., normal shutdown and cooling with steam generator, normal shutdown and cooling with residual heat removal system etc.) is the intrinsic failure probability that is considered in normal cases	Best judgment	BE1

3.2.2. Identification of model parameters affected by the assumption of interest

The model parameters in the PRA model are the probabilities of the basic events in the event tree. As the clogging can lead to the loss of component cooling system (CCS) or the loss of the pumps in the auxiliary feedwater system, the assumption As_3 is related to the two basic events BE3 and BE4, as presented in Table 5.

3.2.3. Assessment of the belief in deviation

Experts from EDF are invited to assess the belief in deviation. In this assumption, the probability that the clogging is not detected and filtered is 1 ($P = 1$), while in reality, the clogging is usually detectable and can be filtered, which means that the true value of this probability is less than 1 ($P < 1$), leading to a lower risk than the value calculated using the assumed model parameters. Therefore, the experts think that this assumption is very conservative, indicating that the assumption deviation might reduce the value of the risk metric.

Some observations can also help the expert to better understand the assumption and evaluate the belief in deviation, as shown in Table 6.

Table 6 Assessment of the belief in deviation

Aspects	Assessment
What could cause the assumption to deviate?	The amount of precipitation can usually be predicted. Hence, if the river flooding is caused by precipitation, then, it can be predicted. Unless it is due to barrier rupture, the river level usually increases gradually and can be seen and noticed easily. If there is heavy precipitation, the operators would pay more attention to the water filters on the river and clean the filters to make sure that the water intake is not clogged.
What are the key drivers of those causes?	The fact that the river level increases is a gradual process. The fact that the operators are able to clean the clogging if it occurs.
Has a similar deviation occurred in the past?	Yes.
What evidence is available for supporting the potential for a deviation?	The feedback reports show that a clogging has occurred before and that operators were able to see it and manage it.

Based on the analysis illustrated in Table 4, the belief in deviation was assigned to be 70%. Therefore, we have $B = 3$.

3.2.4. Evaluate the amount of believed deviation from the true value

Experts in EDF are asked to evaluate, based on their beliefs, the amount of assumption deviation from the true values. The experts have assigned the amount of deviation in percentage to be $D = -50\%$, meaning that the experts believe that the assumption is conservative and deviating towards a higher risk.

3.2.5. Evaluate the strength of knowledge

The strength of knowledge has been evaluated as indicated in Sect. 2.2.6. The strength of knowledge attributes are evaluated separately, as shown in Table 8.

Table 7 Strength of knowledge criteria and weights.

Attribute	Weight	Score
Reasonability and realism of assumptions (k_1)	0.13	1
Availability of reliable data and information (k_2)	0.13	2
Phenomenological understanding (k_3)	0.42	1
Agreement among peers (k_4)	0.16	1
Level of expertise and competence of the experts (k_5)	0.16	2

The overall knowledge score K is calculated using Eq. (3):

$$K = \sum_{i=1}^{5} w_i . k_i = 1.29$$

Then, based on the criteria defined in Sect. 2.2.5, we have $K = 1$.

3.2.6. Determine the context of decision making and define the safety objective

The context of the decision making in this case study is to compare a risk metric to a safety limit. The risk limit for core meltdown varies between 1×10^{-5} and 1×10^{-4} (Knochenhauer & Holmberg 2012).

As the flooding events are usually site-specific (IAEA 2009), the contribution of the external flooding hazard group to core meltdown also varies from one NPP to another. Moreover, we consider only a part of the external flooding PRA model in this case study (through the reduced-order model). Accordingly, for illustration purposes, we artificially set the safety limit of the considered PRA model to be $R_{lim} = 1.6 \times 10^{-8}$.

3.2.7. Identify the margin of deviation

As the assumption As_3 affects the basic events BE_3, BE_4, the vector of basic events' probabilities related to the assumption are $P_m = (p_{BE_3}, p_{BE_4})$. Accordingly, the deviated risk function can be expressed using Eq. (5):

$$R' = R_{th} = R_{lim} = f(p_1, p_2, \acute{p}_{BE_3}, p'_{BE_4}, p_5, \dots p_{10})$$
$$= f(p_1, p_2, (1+\mu)\cdot p_3, (1+\mu)\cdot p_4, p_5 \dots p_{10})$$

The solver in Microsoft Excel is used to solve Eq. (6), with $R_{lim} = 1.603 \times 10^{-8}$. The resulted margin of deviation is $\mu_{As_3} = 26.40\%$. The margins of deviation for the remaining assumptions are calculated in a similar way, as presented in Table 8 next in Sect. 3.2.8.

3.2.8. Evaluate the overall criticality based on the decision flow diagram

As illustrated in Sect. 2, the overall criticality of assumptions deviation is assigned based on the decision flow diagrams in Figure 6-8. For the assumption of interest (As_3), the belief (likelihood) in the deviation is assigned to be 70% (level 3). The difference between the deviation margin and the amount of believed deviation is 76.40%. The strength of knowledge is assessed to be $K = 1$. For an acceptance-criteria decision-context, this means that we believe that we are under the safety limit, and the deviation is not considered critical and can be accepted. On the other hand, our belief is based on weak knowledge, which makes it less credible. Following the decision flow diagram in Figure 6, the criticality of this assumption is $C = 2$. Accordingly, the assumption is not very critical and listed in the "waiting list", which means that it is accepted unless there are other criteria and information on other assumptions deviations that change the evaluation.

The same steps are repeated for each assumption. The scores and the evaluation corresponding to each criterion for each assumption are presented in Table 8 together with their final criticality scores.

Table 8 Assumption-deviation criticality and criticality criteria assessment

A_i	Type	BEs	l_i : B_i	D_i	μ_i	$\Delta\mu_i$	K_i	C_i
1	Conservative	BE2	95%:3	-90%	∞	∞	1	2
2	Best judgment	BE3	30%:2	90%	35.11%	-54.89%	2	1
3	Conservative	BE3, BE4	70%:3	-90%	26.40%	116.40%	1	2
4	Best judgment	BE3, BE4	5%:1	5%	26.40%	21.40%	3	3
5	Best judgment	BE1	50%:3	50%	24.22%	-25.78%	3	1

6	Conservative	BE2-BE10	90%:3	-70%	20.38%	90.38%	1	2
7	Best judgment	BE2-BE10	40%:3	30%	20.38%	-9.62%	2	1
8	Conservative	BE8	20%:1	-30%	869.95%	899.95%	1	2
9	Optimistic	BE1	40%:3	30%	24.22%	-5.78%	2	1
10	Best judgment	BE1	5%:1	5%	24.22%	19.22%	3	3

As shown in Table 8, the different assumptions have three levels of criticality i.e., 1; 2; 3 (very critical; not very critical; not critical). The corresponding actions that need to be taken by decision makers and analysts are respectively:

(i) The deviation is very likely to happen. Besides, the assumption deviation has severe influence on the decision making and might lead to exceedance of the safety limit. Further analysis and justification of the assumption is required.

(ii) The assumption can be trusted to support decision making if the risks of the deviation from other assumptions are all not critical (C=3). Further analysis and justification of the assumption is needed only when multiple other assumptions are also in this state.

(iii) An assumption deviation is unlikely to happen or, if it happens, it does not affect the decision making. The assumption can be trusted and decisions can be made based on the current assumption.

4. Discussion and conclusions

In this paper, we have extended the approach of Khorsandi and Aven (2017) for evaluating assumptions deviations in probabilistic/quantitative risk assessments. The extended framework covers a new context of decision making very relevant in practice, namely, that of comparing alternatives (rather than comparing a single alternative against a safety objective) and an additional type of assumptions, namely, conservative assumptions (rather than just the best judgment type of assumptions). An integrated metric, the criticality of assumption deviation, is defined and evaluated based on the extended framework through the use of decision flow diagrams. The developed framework is applied to a case study of a PRA model of the external flooding hazard group of an NPP. The implementation of the framework has shown its feasibility and its ability to cover different types of assumptions and to provide a more complete evaluation of the assumption deviation.

The use of decision flow diagrams has both pros and cons. The pros are that these diagrams facilitate a standardized assumption deviation risk assessment, increasing both the transparency and efficiency of the assessment. These are desirable attributes in case of peer review of the assessment and considering the large number of assumptions typically involved in PRAs. A con of such diagrams are that they give a “mechanical” assessment procedure where the assessment is based on strict rules rather than the use of overall judgements. Another possible limitation of the current research that need to be addressed in the future is that it analyzes the deviation risk for one assumption at a time and, thus, fails to take into account the deviation risk for several assumptions simultaneously.

Acknowledgements

The work on this article was performed in part while R. Flage was visiting CentraleSupélec in the period February-March 2018. He would like to acknowledge the financial support from CentraleSupélec, as well as his co-author Professor E. Zio for making his stay possible, and memorable.

References

1. Aven, T. (2013). Practical implications of the new risk perspectives. Reliability Engineering & System Safety, 115, 136-145.
2. Bani-mustafa, T. et al., 2018. Strength of Knowledge Assessment for Risk Informed Decision Making. In *Esrel*. Trondheim.
3. EPRI, 2015. *An Approach to Risk Aggregation for RisK-Informed Decision-Making*, Palo Alto, California.
4. EPRI, 2006. *Guideline for the Treatment of Uncertainty in Risk-Informed Applications: Applications Guide*, Palo Alto, California: 1013491.
5. Goerlandt, F. & Montewka, J., 2014. Expressing and communicating uncertainty and bias in relation to Quantitative Risk Analysis. *Safety and Reliability: Methodology and Applications*, 2(13), pp.1691–1699. Available at: http://www.crcnetbase.com/doi/abs/10.1201/b17399-230.
6. IAEA, 2003. *External Events Excluding Earthquakes in the Design of Nuclear Power Plants*,
7. IAEA, 2011. IAEA-Publication8635.
8. IAEA, 2009. *Meteorological and Hydrological Hazards in Site Evaluation for Nuclear Installations*,
9. Khorsandi, J. & Aven, T., 2017. Incorporating assumption deviation risk in quantitative risk assessments: A semi-quantitative approach. *Reliability Engineering & System Safety*, 163, pp.22–32.
10. Knochenhauer, M. & Holmberg, J.E., 2012. Guidance for the definition and application of probabilistic safety criteria. *Proceedings of PSAM 10 International Probabilistic Safety Assessment & Management*.
11. Saaty, T.L., 2008. Decision making with the analytic hierarchy process. *International Journal of Services Sciences*, 1(1), p.83.
12. Suter II, G. W. (2007). Ecological risk assessment. Second Edition. Taylor & Francis.

Appendix IV (P4):

Tasneem Bani-Mustafa, Zhiguo Zeng, Enrico Zio and Dominique Vasseur “A practical approach for evaluating the strength of knowledge supporting risk assessment models” Safety Science (Under review)

A practical approach for evaluating the strength of knowledge supporting risk assessment models

Tasneem Bani-Mustafa [(1)], Zhiguo Zeng [(1)], Enrico Zio [(1) (2)], Dominique Vasseur [(3)]

[(1)] *Chair on System Science and the Energetic Challenge, EDF Foundation*

Laboratoire Genie Industriel, CentraleSupélec, Université Paris-Saclay,

3 Rue Joliot Curie, 91190 Gif-sur-Yvette, France

[(2)] *Energy Department, Politecnico di Milano, Via Giuseppe La Masa 34, Milan, 20156, Italy*

[(3)] *EDF R&D, PERICLES (Performance et prévention des Risques Industriels du parc par la simulation et les Etudes) EDF Lab Paris Saclay - 7 Bd Gaspard Monge, 91120 Palaiseau, France*

Abstract

In this paper, we develop a new quantitative method to assess the Strength of Knowledge (SoK) of a risk assessment. A hierarchical framework is first developed to conceptually represent the SoK in terms of three attributes (assumptions, data, phenomenological understanding), which are further broken down in sub-attributes and "leaf" attributes to facilitate their assessment in practice. The hierarchical framework, is, then, quantified in a top-down bottom-up fashion for assessing the SoK. In the top-down phase, a reduced-order risk model is constructed to limit the complexity and number of basic elements considered in the SoK assessment. In the bottom-up phase, the SoK of each basic element in the reduced-order risk model is assessed based on predefined scoring guidelines and, then, aggregated using a weighted average of "leaf" attributes , where the weights are determined based on the Analytical Hierarchical Process (AHP). The strength of knowledge of the basic events is in turn, aggregated using a weighted average to obtain the SoK for the whole risk assessment model. The developed methods are applied to a real-world case study, where the SoK of the Probabilistic Risk Assessment (PRA) models of a Nuclear Power Plants (NPP) is assessed for two hazards groups, i.e. external flooding and internal events.

Keywords

Strength of Knowledge (SoK), Probabilistic Risk Assessment (PRA), Risk-Informed Decision Making (RIDM), Multi-Hazards Risk Aggregation (MHRA), Event Tree (ET), Nuclear Power Plant (NPP).

Acronyms

AFW: Auxiliary feedwater

AHP: Analytical Hierarchy Process

BE: Basic Events

CDF: Core-Damage Frequency

DAMA: Data Management Association's

DIKW: Data-Information-Knowledge-Wisdom

EDF: Electricité De France

EUROSTAT: EUROPEAN STATISTICS

GAGAS: Generally Accepted Government Auditing Standards

IAEA: International Atomic Energy Agency

IE: Initiating Events

LOCAs: Loss of Coolant Accidents

MCSs: Minimal Cut Sets

MHRA: Multi-Hazards Risk Aggregation

NPP: Nuclear Power Plants

NS/SG: Normal Shutdown with cooling using Steam Generator

NUSAP: Numeral Unit Spread Assessment Pedigree

PRA: Probabilistic Risk Assessment

QRA: Quantitative Risk Assessment

RIDM: Risk-Informed Decision Making

SoK: Strength of Knowledge

1. Introduction

In PRA, models are developed to calculate some probabilistic indexes for risk characterization (Flage & Aven 2009). These probabilistic indexes, believed to be objective but with unknown values, express the irreducible "aleatory uncertainty" in the related systems and processes (Helton & Burmaster 1996), (Helton *et al.,* 2004), (Flage & Aven 2009). However, since these indexes are calculated by the developed "model of the world" (Apostolakis 1990), they are conditioned on the knowledge on the problem. Lack of knowledge will result in additional uncertainty in the PRA results, known as "epistemic uncertainty" (Helton & Burmaster 1996), (Helton *et al.,* 2004), (Flage & Aven 2009). It is well-accepted in the risk assessment community that epistemic uncertainty needs to be properly quantified and taken into account in PRA. Since epistemic uncertainty depends on the Strength of Knowledge (SoK), quantifying the knowledge that supports risk modeling and assessment is an indispensable task in probabilistic risk assessment (PRA) (Askeland *et al.,* 2017), (Aven 2017b). In fact, some experts even propose to use "uncertainty", instead of "probability", as a main component of risk and interpret the probability as knowledge-based expressions of uncertainty (Flage & Aven 2009), (Aven 2013a), (Aven 2013b), (Aven & Krohn 2014). Beyond that, other experts insist on using the term "characterizing" rather than "measuring" when talking about risk metrics like the Core-Damage Frequency (CDF), in order to highlight the belief that the metrics obtained from PRA models provide only a representation of the state of knowledge (EPRI 2015).

However, the existing works on epistemic uncertainty quantification and propagation (for example, including but not limited to subjective probability, imprecise probability, evidence theory, possibility theory, etc.) aim at developing mathematical frameworks to represent the epistemic uncertainty in the input and then propagate the uncertainty to quantify the epistemic uncertainty in the output. For example, in imprecise probability, the epistemic uncertainty is represented using probability intervals and propagated following the rules of probability theory. How to determine the probability intervals for the input parameters, however, is not fully addressed in these methods. With respect to this problem, the assessment of SoK is a critical step, as the epistemic uncertainty is directly related to the SoK. In fact, quantifying the SoK is even more important in risk-informed decision making. For example, in the current multi-hazards risk aggregation methods, the aggregation is done by a simple arithmetic summation of risk from different contributors and the final results are compared to quantitative safety goals and acceptance criteria to support decision making. However, this simple arithmetic summation does not take into account the fact that the risk estimates from different contributors are based on different degrees of knowledge and therefore, might have different degrees of realism (EPRI 2015). Another example is that when the decision maker needs to choose among different alternatives based on the estimated risk, simply choosing the alternative with a lower risk estimate without considering the degree of knowledge might not be the right choice.

SoK of a risk assessment model refers to the level of knowledge that supports the model. It affects the trust one has on the results obtained by the risk assessment and the decisions that are based on them (Aven

2013b), (Bani-Mustafa *et al.*, 2017b). For example, in the risk assessment of Nuclear Power Plants (NPPs), the SoK of an external flooding risk model may be relatively low, due to the fact that the phenomena involved are not so well-understood and the data are limited: then, it is expected that conservative decisions would be taken even if the risk assessments were to yield optimistic results (EPRI 2015). The importance of considering SoK in risk assessment has led researchers to formulate frameworks in which risk is described not only by traditional elements (like scenarios, likelihoods and consequences (Aven 2012)), but also by elements directly related to knowledge (Montewka et al. 2014), (Aven 2012), (Aven & Ylönen 2016), (Aven 2013b), (Bjerga & Aven 2015). For example, in the Data-Information-Knowledge-Wisdom (DIKW) hierarchy in (Aven 2013a): the SoK is explicated to complement the two traditional risk dimensions of consequence and uncertainty (Aven 2017b).

Only very few works, however, directly address the issue of how to evaluate the SoK of a risk assessment model. A semi-quantitative approach for evaluating the SoK is proposed by Goerlandt and Montewka (2014), based on four criteria: (i) phenomenological understanding and availability of trustable predicting models; (ii) reasonability and realism of assumptions; (iii) availability of reliable and relevant data and information; (iv) agreement/disagreement among peers. Three levels of SoK are identified based on the degree that the previous criteria are satisfied. Aven (2013b) considers the SoK that supports the determination of probability intervals used in Norway national risk assessment (NRA) and a risk analysis concerning a Liquefied Natural Gas (LNG) plant. In Aven and Ylönen (2016), safety regulations of the oil & gas and nuclear industries have been enhanced by assessing the SoK which probabilities of risk acceptance criteria are based on. Bjerga and Aven (2015) develop an adaptive risk management plan for the oil and gas industry, where the SoK that supports the estimation of probability intervals is assessed and represented as an additional dimension of a risk matrix. In Montewka *et al.* (2014a), a qualitative description of uncertainty in maritime-based risk analysis and decision making is presented by developing a two-dimensional scoring system taking into account the SoK. Berner and Flage (2016) consider the risk assessment of lifting riserless light well intervention equipment on the Norwegian continental shelf and assess the SoK on which important assumptions of risk assessment are based. Askeland *et al.* (2017) adapt the assessment framework in Flage and Aven (2009) and apply it on security risk assessment, where a fifth criterion, i.e., knowledge scrutinization, is added to the four criteria defined by Flage and Aven (2009) for SoK assessment. The SoK is, in turn, classified into three levels, i.e. weak, strong and medium (Askeland *et al.*, 2017). More examples of the SoK evaluation of the risk assessment models by semi-quantitative models can be found in (Abrahamsen *et al.*, 2016), (Aven 2017a), (Berner and Flage, 2016), (Khorsandi & Aven 2017), (Haouzi et al. 2013).

Another method proposed for SoK assessment is the assumption deviation risk method, whose standpoint is that poor assumptions are main sources of weak knowledge and, hence, efforts should be made for evaluating the solidity of assumptions on which risk analysis is based (Aven 2013b); (Berner and Flage, 2016). The method identifies the criticality of assumptions by assigning crude risk scores for the main assumptions of the risk assessment model, which cover: (i) the possible deviation from the assumptions

and the associated consequences; (ii) the uncertainty of this deviation; (iii) the background knowledge that supports the assumptions. Similarly, Berner and Flage (2016) define guidelines to treat the uncertainty associated with six typical settings that correspond to different levels of assumptions deviations. In addition to this method, Berner and Flage (2016) identifies three other approaches for treating uncertain assumptions: (i) law of total expectation; (ii) interval probability; (iii) crude SoK and sensitivity categorization. In the law of total expectation method works for scenarios with strong knowledge and historical data where, a probability distribution is introduced to express the belief on different assumptions. In the case of weak knowledge, on the other hand, interval probability technique can be applied, where the assessors are asked to assign the minimum and maximum values of assumptions and their corresponding believed probability. In the crude SoK and sensitivity categorization method, the criticality of assumption is assessed by assessing the strength of knowledge on which the assumptions are made, as well as the dependency of risk assessment on this assumption.

Goerlandt and Reniers (2016) propose to assess and visualize uncertainty in risk assessment through probability-consequence diagrams, in which the assumption deviation risk is visualized along with a segmented strength-of-evidence assessment. Khorsandi and Aven (2017) emphasize the importance of integrating the assumption deviation risk in quantitative risk assessment in order to provide a complete representation of the risk and apply the method to a case study from the offshore industry. Aven (2017b) suggests using the assumption deviation risk method as a complement to the quantitative risk assessment, to improve traceability of the results and perform a more responsible RIDM.

As seen from the above, most of the existing methods are qualitative in nature, wherein the assessment is done based on some crudely defined scoring criteria, which limits the practical application. In practice, however, a quantitative evaluation of SoK is needed for operationally supporting RIDM. Also, many SoK attributes are difficult to evaluate directly and, yet, their evaluation is carried out directly by simple scoring based on a plain description of the attributes, which can be difficult and imprecise in practice. To make a quantitative evaluation feasible, the high-level attributes need to be broken down into more tangible sub-attributes. Besides, the SoK cannot be evaluated directly on the entire risk assessment model: rather, a feasible approach should consider the SoK of the basic and most relevant elements. Compared to the existing methods, the contributions of this paper include: (i) A hierarchical framework is developed to conceptually represent the SoK and break it down into tangible sub-attributes and "leaf" attributes to facilitate the assessment in practice; (ii) Detailed scoring guidelines are given for evaluating the bottom-level attributes in the SoK assessment framework; (iii) A top-down bottom-up approach is developed for the practical evaluation of the SoK supporting the PRA model. More specifically, the work in this paper is rather an attempt to support RIDM by "measuring what we know instead of what we don't know". This work is directed towards supporting risk-based decision making by giving indices on the state of knowledge on which the risk assessment is based. Hence, the main goal of this paper is to develop a framework that measures practically the concept of "strength of knowledge" that has been introduced recently by some colleagues and accepted and used by others for supporting the risk assessment (Milazzo

& Aven 2012), (Aven 2013b), (Montewka et al., 2014), (Goerlandt & Montewka 2015), (Valdez Banda et al. 2015), (Berner & Flage 2016a), (Berner & Flage 2016b), (Goerlandt & Reniers 2016). The paper aims to complement and formulate in a practical way the previous attempts developed for evaluating the SoK supporting the RIDM.

However, it should be noted that although SoK is an important contributor to the trust in the PRA results, it is not the only contributor. Other factors, e.g., the quality of the modeling process, also need to be considered if one wants a complete evaluation of the PRA trustworthiness. The current work focuses on the SoK, i.e., how much we know about the system and processes related to risk. The specific focus is on complementing and formulating, in a practical way, the previous attempts for evaluating the SoK supporting the RIDM (Milazzo & Aven 2012), (Aven 2013b), (Montewka *et al.*, 2014), (Goerlandt & Montewka 2015), (Valdez Banda et al. 2015), (Goerlandt & Reniers 2016), (Berner & Flage 2016a), (Berner and Flage, 2016).

In this paper, we propose a quantitative assessment of SoK. A hierarchical framework is developed in Section 2 to conceptually describe SoK and relate it to its major contributors. The framework is, then, developed into a top-down and bottom-up method for SoK assessment (Section 3), considering the essential constituents of the risk assessment model. In Section 4, a case study of two hazard-group in Probabilistic Risk Assessment (PRA) models of a Nuclear Power Plant (NPP) is presented. Finally, the paper is concluded in Section 5 with a discussion.

2. A hierarchical framework for SoK assessment

In this section, we construct a conceptual framework to describe the SoK that supports a PRA. The main attributes that contribute to the SoK are identified from the literature and organized hierarchically based on the framework proposed in Flage and Aven (2009), but adjusted and expanded to include more contributors and facilitate the practical implementations. In Sect 2.1, we illustrate the development of the framework. In Section 2.2, we formally present the framework and define its attributes.

2.1 Framework development

In this section, we survey the attributes typically considered in existing works for SoK assessment and argue the importance of including specific criteria in defining the strength of knowledge and finally, organize them in a hierarchical framework for practical assessment.

Let's take the PRA models as an example to illustrate our arguments. Different steps need to be followed to construct and operate correctly a PRA model, as shown in Figure 1 (Stamatelatos et al. 2011), (NRC 1983).

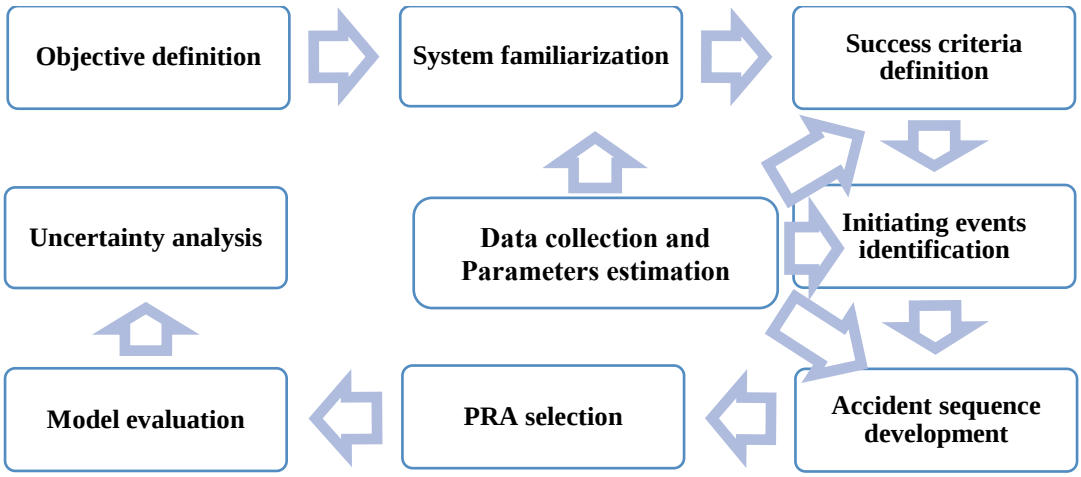


Figure 1 Typical PRA process flow (Stamatelatos et al. 2011), (NRC 1983).

Now, let's take each step and elicit the different knowledge required for successfully implementing each step. The required knowledge is summarized in Table 1. Please note that since we are not concerned about the quality of the analysis in this work, some steps in Figure 1 are not relevant and, therefore, not considered in Table 1, such as model evaluation, PRA selection, etc.

Table 1 PRA's typical steps requirements

Objective	Requirements for achieving the objectives *(required knowledge)*
Objectives definition: The defined objectives need to be unambiguous and clearly defined and understood by the risk analyst	• The objectives are defined based on widely accepted quality standards for implementing PRA • Sufficient data and information are available to support the definition of the objectives *(Explicit knowledge, in forms of data, information and understanding)* • Availability of experts who have sufficient experience in the domain and low value-ladenness and are able to elicit unexpected and unexperienced hazards leading to initiating events *(implicit knowledge in forms of phenomenological understanding provided by reliable experts with low value ladenness)*
System familiarization: The analysts need to be familiar with system structure and understand the functional principle	• The technology of the systems is very mature and the functional principles of the system are well-understood *(explicit and implicit knowledge in the form of phenomenological understanding)* • There are abundant design and operation manuals to support the analysis *(explicit knowledge in forms of data and industrial evidence)* • Availability of experts who have sufficient experience in the domain understanding of the problem and the related systems, and low value-ladenness *(implicit knowledge in forms of phenomenological understanding provided by reliable experts with low value ladenness)*
Success criteria definition: All the possible success and failure criteria of the missions and systems need to be identified and clearly defined	• There are abundant technical reports that allow the understanding of different the systems and the backup systems *(explicit knowledge in forms of data and phenomenological understanding)* • There is abundant detailed past experience operation, transient, incidents and accident reports *(explicit knowledge in forms of data and phenomenological understanding)* • The analysts have access to related technical reports and a good understanding of functional principles of the system *(explicit knowledge in forms of data and explicit and implicit in forms of phenomenological understanding)* • The availability of experts who have sufficient experience and low value-ladenness *(implicit knowledge in forms of phenomenological understanding and solid assumptions provided by reliable experts with low value ladenness)*

Initiating events identification: All possible events that might lead to an abnormal operation or to an accident should be clearly defined	• There are abundant detailed past experience reports about different initiating events *(explicit knowledge in forms of data)* • The analysts have a good understanding of the interconnections between systems and the dependency on system failures *(implicit knowledge in forms of phenomenological understanding)* • The analysts have access to related technical reports and a good understanding of functional principles of the system *(explicit knowledge in forms of data and explicit and implicit in forms of phenomenological understanding)* • The process of identifying initiating events follows well-accepted quality control guidelines for PRA • Availability of experts who are able to elicit unexpected and unexperienced hazards leading to initiating events *(implicit knowledge in forms of phenomenological understanding)* • The completeness of the identification process is verified by peer review of qualified experts *(implicit knowledge in form of agreement among experts)* • The availability of experts who have sufficient experience and low value-ladenness (*implicit knowledge in forms of phenomenological understanding provided by reliable experts with low value ladenness)*
Accident sequence development: The possible abnormal-operation progressions are well understood and clearly defined, and cover all the possible scenarios	• The evolution sequence is known and well represented *(explicit and implicit knowledge in forms of phenomenological understanding)* • The functional principles of the system are well-understood *(explicit and implicit knowledge in forms of phenomenological understanding)* • The environment and phenomena surrounding and that might affect the system are well-understood *(explicit and implicit knowledge in forms of phenomenological understanding)* • The availability of detailed abnormal activities reports that allow understanding the sequential development of an activity (*explicit knowledge in forms data)* • The availability of experts with sufficient experience that allow developing thoroughly the different scenarios of any abnormal activity (*implicit knowledge in forms of phenomenological understanding and solid assumptions provided by reliable experts with low value ladenness)*

Data collection and parameters estimation: The data needed for parameters estimation and model evaluation are complete and clearly represented	• The operation, maintenance, and failure reports are available *(Explicit knowledge in from of data)* • The abundance of highly reliable data for the estimation of input parameters *(Explicit knowledge in from of reliable data)* • Availability of credible models to calculate the model parameters • The process of data collection and representation follows quality control guidelines that ensure its reliability and quality (*Explicit knowledge in from of reliable data)*

It can be seen from Table 1 that two forms of knowledge appear in PRA: explicit knowledge, which refers to all types of knowledge that can be explicitly transferred, including data, documented established theory and explanation of phenomena and any kind of undocumented but transferable data, information and phenomenological understanding; and the implicit knowledge that is owned by the individuals to support the risk assessment but cannot be transferred (Davies 2001). The knowledge in Table 1 can also be categorized into four aspects: "data" for input parameters, hazards, initiating events and accidents sequences; "understanding of phenomena" related to the function of the systems, their interrelations, and the surrounding environment; "expert's past experience and knowledge" that allow predicting the inexperienced hazards, unknown parameters and "assumptions" regarding the development of the scenarios and construction of the model.

In fact, the four aspects, i.e., data, understanding of phenomena, expert experience and assumptions have long been considered in the literature as the main contributors to the SoK. For example, Nowakowski *et al.,* (2014) argue that unlike the traditional Greek perspectives of knowledge as being justified true belief, the risk analysis propositions are in the form of assumptions and phenomenological understanding shaped by history (data) and present. Also, a well-accepted conceptual framework was defined by Flage and Aven (2009) comprised of four components: the inter-alia assumptions and presuppositions (solidity of assumptions), historical field data (availability of reliable data), understanding of phenomena and agreement among experts. However, since the "agreement among experts" are more related to the construction of the model and making assumptions (either assumptions on model structure or assumptions on parameter values), it is considered in this work as a sub-attribute of the "solidity of assumptions" and extended to cover further value-ladenness of the assessors. The first three components in (Flage and Aven, 2009) are, then, adopted as the top-level attributes of our conceptual hierarchical framework for SoK. In the following subsections, we elaborate on these three attributes by surveying their contributing elements one by one.

2.1.1 Solidity of assumptions

In risk analyses, assumptions are inevitably made by experts because of incomplete knowledge, data, information and understanding of the phenomena involved, for simplifying the analysis when necessary (Kloprogge *et al.,* 2011). These assumptions might be in different forms, such as assumptions made by

experts about the values of input parameters, the environmental conditions surrounding the system of interest, the scenarios, and consequences in a model. In fact, the assumptions considered can be understood as related to any kind of input or conditions that are assumed and acknowledged to possibly deviate from reality (Berner and Flage, 2016). Such assumptions are part of the background knowledge that supports the analysis. Simple assumptions compose a source of uncertainty "hidden in the background knowledge" of the risk assessment (Berner and Flage, 2016). The SoK that supports risk assessment, therefore, depends on the solidity of the assumptions made (Boone et al. 2010).

Few methods have been proposed for evaluating the quality of assumptions and treating the uncertain assumptions in risk assessment. Numeral Unit Spread Assessment Pedigree (NUSAP) is proposed to directly assess the quality of assumptions for complex problems (Van Der Sluijs et al. 2005), (Boone et al. 2010), (Kloprogge *et al.*, 2011), (De Jong *et al.*, 2012). This method allows analyzing the strength, importance and potential value-ladenness of assumptions through a pedigree diagram. The pedigree allows the evaluation of assumptions given seven criteria: (i) plausibility; (ii) inter-subjectivity peers; (iii) inter-subjectivity stakeholders; (iv) choice space; (v) influence situational limitations; (vi) sensitivity to view and interests of the analyst (vii) and influence on results. Three scores are defined in the pedigree, ranging from zero to two (0-2); each, one correspond to a degree of fulfillment of the criterion. The scheme covers clearly some social and value-ladenness aspects affecting the assumptions, as well as their implication on the results (Van Der Sluijs et al. 2005), (Boone et al. 2010), (Kloprogge *et al.*, 2011), (De Jong *et al.*, 2012). However, it does not cover explicitly the subjectivity and knowledge of the experts who make the assumptions. In Zio (1996) various criteria are defined for evaluating the value-ladenness and confidence in experts' judgments, such as the source of information, the degree of non-biasedness, the degree of independence, and the personal interests etc. These factors should also be considered when evaluating the solidity of assumptions.

We group the aforementioned contributing factors into three categories, i.e. quality (solidity) of assumptions, the sensitivity of assumptions and value-ladenness. Quality (solidity) of assumptions refers to the degree to which the assumptions are realistic and reasonable and affects greatly the solidity of assumptions and their effectiveness in supporting the risk assessment (Berner and Flage, 2016). Value ladenness refers to the degree of the inevitable bias by the assessors who make the assumptions, due to their subjectivity, personal perceptions, external limitations, etc. (Zio 1996), (Kloprogge et al., 2011). This attribute is directly connected to the quality of assumptions, since they are made by the assessor. It might be argued that the value-ladenness affect other attributes of the strength of knowledge, as the other attributes are in form of explicit knowledge that can be documented and transferred "objectively" without being affected by the expert's subjectivity, unlike the "assumptions" that are made based on expert's judgment and greatly affected by subjectivity. Finally, the sensitivity of assumptions considers the degree to which the models' output varies if the assumptions are changed into the alternative ones (Stirling 1999), (Saltelli et al. 2013). Hence, it is related to the model output and not the strength of knowledge supporting the model input. Therefore, it is not considered in our developed framework. In particular, the value-

ladenness is further expanded into seven sub-attributes to cover the most important factors that affect the expert's judgment (Zio 1996): (i) the personal knowledge; (ii) the sources of information; (iii) the non-biasedness; (iv) the relative independence; (v) the past experience; (vi) the performance measure; (vii) the agreement among peers. Detailed descriptions of these attributes can be found in Section 2.2.

2.1.2 Availability of reliable data

Data is considered the bottom tier of the DIKW hierarchy as defined in (Hey 2004), (Aven 2013a). When processed, data yield information that becomes knowledge when combined with experience and judgment (Kidwell *et al.,* 2000), (Rowley & Hartley 2017). Thence, the amount of data available is a natural measure of the strength of knowledge. However, having a large amount of data alone does not necessarily indicates strong knowledge, as the available data might be of low quality. Some expert might prefer few data of high reliability over large amount of data of low reliability. In other words, the reliability of data is also very important for supporting PRA. In Flage and Aven (2009), apart from the availability of data, the reliability of data is also identified as an essential element for evaluating the SoK. Hence, both availability and reliability of data are considered in the developed framework for SoK assessment, as shown in Figure 2.

Data availability can be assessed qualitatively. For example, Flage and Aven (2009) quantify the degree of the availability of data verbally: data are not available, much data are available etc. Data availability can also be quantified quantitatively by numerical indicators related to the amount of data. For example, failure data are collected from different components and over various time intervals: the data collection time interval and the number of components from which the data is collected, can, then, be regarded as numerical indicators of data availability.

Data reliability refers to the representativeness of the data in the context of the purpose that they are used for (Morgan & Waring 2004). Various attributes have been defined in the literature for evaluating data reliability. For example, in computer science, data reliability is evaluated by its completeness, accuracy, and consistency (Roth 2009). Tests are made to verify whether the data meet the "Generally Accepted Government Auditing Standards" (GAGAS), with respect to three aspects:

(i) Sufficiency: referring to the "*completeness*" of the data in the context of supporting the finding.
(ii) Competence: referring to the closeness of data to reality ("*accuracy*") and also the *validity*, *completeness,* and *non-alteration* of data.
(iii) Relevance: referring to the logical and sensible relationship of the data to the finding it supports ("*consistency*"), as well as the age of the data ("*timeliness*").

A survey of 39 articles conducted by Chen *et al.* (2014) identifies main attributes of data reliability (referred as data quality in their paper) as completeness, accuracy, timeliness, validity, periodicity, relevance, reliability, precision, integrity, confidentiality, etc. Among them, completeness, accuracy, and timeliness have been most frequently used in testing data reliability (Chen et al. 2014). To assess the reliability of statistical data, EUROPEAN STATISTICS (EUROSTAT) recommends six attributes, i.e., relevance, accuracy, timeliness, comparability, coherence, accessibility and clarity (Bergdahl et al. 2007). International Atomic Energy Agency (IAEA) identifies relevance, timeliness, accuracy, and completeness

as main attributes for data reliability in the nuclear industry (IAEA 1991). Six attributes, i.e., completeness, uniqueness, timeliness, validity, accuracy, consistency, are recommended in the Data Management Association's (DAMA) white paper for evaluating data reliability (DAMA 2013).

In general, choosing different data reliability attributes is an organization and context-wise task (DAMA 2013). In this paper, we identify the following five attributes for assessing data reliability, based on the literature review above and their relevance to the SoK of risk assessment: (i) completeness; (ii) timeliness; (iii) validity; (iv) accuracy; (v) consistency and relevance. Most of these attributes are considered by different organizations due to their importance (IAEA 1991), (Bergdahl et al. 2007), (DAMA 2013). The completeness of data is obviously a very important issue to ensure that the data can fulfill its purpose and do not cause misleading. The timeliness guarantees that the data are up to date and keep up with the development in the technology and the measuring techniques. The validity ensures that data are collected and stored in a managed and standardized way to keep its integrity and facilitate access without errors. The accuracy of data ensures that the data are of value in representing reality and do not lead to misinformation. Finally, the consistency and relevance of data are very important to ensure that they are collected from relevant and consistent sources in a way that is suitable for the desired purpose. Detailed descriptions of these attributes can be found in Section 2.2.

2.1.3 Understanding of phenomena

In this study, understanding of phenomena refers to the comprehension of the events, phenomena and system's functionality that are involved in the risk modeling and assessment. The more the phenomena are understood, the more knowledge for supporting the risk assessment. As illustrated before, knowledge in risk analysis is characterized in the form of assumptions and phenomenological understanding shaped by history and present to predict the future (Nowakowski et al. 2014). Phenomenological understanding has been identified by many researchers as an important constituent of SoK that is needed to support risk assessment (Flage & Aven 2009), (Goerlandt & Montewka 2014), (Nowakowski et al. 2014). However, few existing works have focused on its assessment. For example, Flage and Aven (2009) evaluate it crudely by introducing verbal expressions such as "not well understood", "well understood", "not available", "much available" etc. However, this kind of evaluation seems very crude since it doesn't overcome the intangibility of this attribute. The attribute itself is intangible and difficult to be evaluated directly without breaking it down to more tangible attributes.

In general, a comprehensive understanding of a phenomenon requires a correct and complete explanation of it (Kelp 2015). So, having a documented explanation of the phenomena, phenomenon-related application experience and abundant experts in the related field can help to understand the phenomenon. This means that the experience gained related to a given phenomenon, the documented pieces of evidence, the application related to the phenomena and the understanding gained by individuals can be indications on the understanding of phenomena. Accordingly, we propose four sub-attributes to evaluate the level of phenomenological understanding: (i) number of industrial evidence; (ii) number of academic evidence;

(iii) number of experts involved; (iv) number of years of experience in the domain. A detailed description of these sub-attributes can be found in Sect 2.2.

2.2 The developed framework

In this section, we present the framework developed, based on the review in Section 2.1. As shown in Figure 2, the SoK, denoted by K (Level 1), represents the solidity of background knowledge that supports a risk model. A high value of K indicates that the model is well supported and, therefore, its results are trustable. The SoK is characterized by three level-2 attributes: solidity of assumptions (A), availability and reliability of data (D), and understanding of the phenomena (Ph). The attribute A measures the plausibility, objectivity and sensitivity of the assumptions upon which the model is based; D measures the amount and reliability of data that support the model evaluation; and Ph measures the degree of comprehension of the phenomena involved in the risk assessment.

The three attributes of level-2 are further decomposed into sub-attributes (Levels 3 and 4) to assist their evaluation in practice. Please note that the breaking-down is designed in such a way that the sub-attributes in the same level of the hierarchy are independent and mutually exclusive. Detailed definitions of the attributes are given in Table 2 and Table 3. Detailed guidelines for the evaluation of the attributes at the bottom levels of the framework are defined in Appendices A-C.

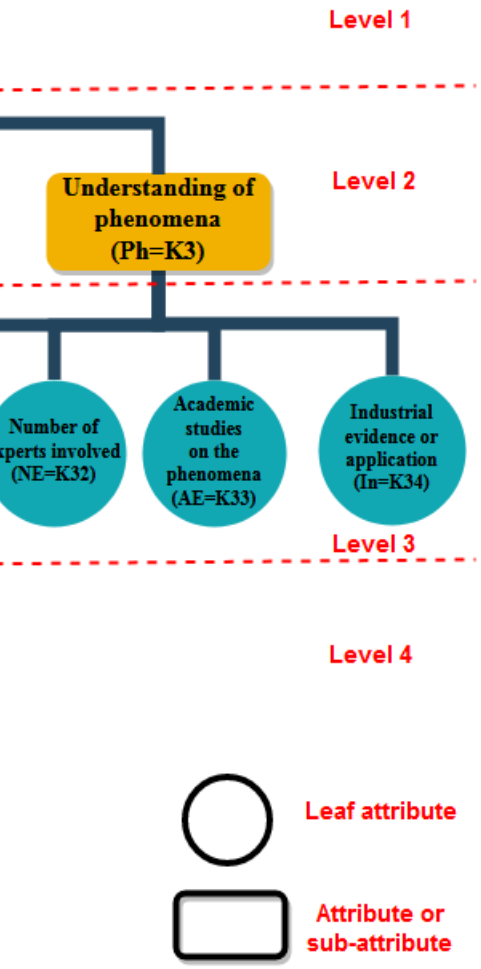


Figure 2 A hierarchical conceptual framework for knowledge assessment

Table 2 Definition of SoK attributes (Level 3)

Attribute	Definition
Value ladenness of the analyst ($VL = K_{12}$)	The degree to which the presumed values and beliefs that are taken as facts, and the assumptions made by experts are affected by the personal points of view, bias, subjectivity, and external or personal limitations
The sensitivity of assumption ($S = K_{13}$)	The degree to which the models' output varies with assumptions
Amount of available data ($AD = K_{21}$)	The quantity of data that supports the modeling and analysis
Reliability of data ($RD = K_{22}$)	The degree to which the available data is complete, accurate and error-free, consistent, valid and representative of reality
Years of experience ($YE = K_{31}$)	The amount of experience (measured in years) regarding a specific phenomenon
Number of experts involved ($NE = K_{32}$)	The number of experts who are explicitly or implicitly involved in understanding the phenomena and the risk analysis
Academic studies on the phenomena ($AE = K_{33}$)	The number of academic resources, i.e., articles, books, etc., available in relation to the phenomena of interest
Industrial evidence and applications on the phenomena ($IE = K_{34}$)	The number of industrial applications and reports related to the specific phenomena or events of interest

Table 3 Definition of SoK attributes (Level 4)

Attribute	Definition
Personal knowledge ($PK = K_{121}$)	The level of analysts' knowledge and relevance to the problem
Source of information ($SI = K_{122}$)	The degree of solidity, relevance, and confidence of the experts' source of information and knowledge
Unbiasedness and plausibility ($U = K_{123}$)	The experts' degree of objectivity and unbiasedness towards personal interest, or an intentional or non-intentional tendency towards a specific subject in the analysis
Relative independence ($RI = K_{124}$)	The degree of independence of the analysts from limitations or external pressures
Past experience ($PE = K_{125}$)	The experts' degree of experience in the related domain and more specifically, in the specific problem under analysis
Performance measures ($PM = K_{126}$)	The experts' degree of professionalism, skills, and competencies, past fulfillment of assigned missions and level of achievement
Agreement among peers ($P = K_{127}$)	The degree to which the assumptions made by different experts are consistent

Completeness ($C = K_{221}$)	The degree to which the collected data contains the needed information for the risk modeling and assessment
Consistency ($Co = K_{222}$)	The degree of homogeneity of data from different data sources
Validity ($V = K_{223}$)	The degree to which the data are collected from a standard collection process and satisfy the syntax of its definition (documentation related)
Accuracy and conformity ($Ac = K_{224}$)	The degree to which data correctly reflects the reality about an object or event
Timeliness ($T = K_{225}$)	The degree to which data are up-to-date and represent reality for the required point in time

3. A top-down bottom-up method for SoK assessment

In this section, we present a top-down bottom-up method to facilitate the practical implementation of the framework proposed in Figure 2 for the evaluation of the SoK supporting risk assessment models. In Section 3.1, we give an overview of the SoK assessment method. In Section 3.2, we show how to break down the risk model into the basic elements of a reduced-order model. Section 3.3 presents the evaluation of relative importance (weights) of SoK attributes using pairwise comparison matrices of Analytical Hierarchy Process (AHP) (Saaty 2008). Finally, in Section 3.4, we illustrate how to aggregate the SoK of the basic elements to evaluate the SoK of the total risk assessment model.

3.1 Procedural steps of the top-down bottom-up method

For the purpose of illustration, we consider the Probabilistic Risk Assessment (PRA) models used in the nuclear industry. Specifically, we refer to the widely applied event tree models. The events probabilities in the event tree model are calculated by fault tree models. The risk index considered is the probability of occurrence of a given consequence (e.g. the probability of core damage in a NPP). For each combination of operation state and scenario, a dedicated risk assessment model (in this case, an event tree) is developed and the total risk index is calculated by summing the values of the risk indexes calculated for each individual risk model:

$$R = \sum_{i=1}^{n_O} \sum_{j=1}^{n_{S,i}} R_{i,j}, \quad (1)$$

where n_O is the number of operation states (O), $n_{S,i}$ is the number of accident sequences (scenarios, S) that are considered in operation state i and can lead to the given consequence of interest. Each $R_{i,j}$ in Eq. (1) quantifies the risk contribution specific to scenario j (e.g., medium flood level) in operation state i (e.g., emergency shutdown).

The risk models for calculating the specific risk index contribution $R_{i,j}$ are characterized by initiating events (IEs), basic events (BEs) and their combinations in minimal cut sets (MCSs). Please note that the initiating events in the PRA model are basic events that trigger the abnormal activity, so it will be treated hereafter as a basic event. Taking the rare-event approximation, $R_{i,j}$ can be calculated by (Zio 2007):

$$R_{i,j} = \sum_{k=1}^{n_{MCS,i,j}} \prod_{q \in MCS_k} P_{BE,q}, \quad (2)$$

where $n_{MCS,i,j}$ is the number of minimal cut sets in the risk model for operation state i and scenario j, MCS_k is the k-th minimal cutset and $P_{BE,q}$ is the occurrence probability of the q-th basic event in MCS_k.

For the following illustration of the SoK assessment procedure, it can be considered that the four elements O, S, MCS and BE fully define the PRA model, as shown in Figure 3. We refer to these four elements as the "constituting elements" of the model.

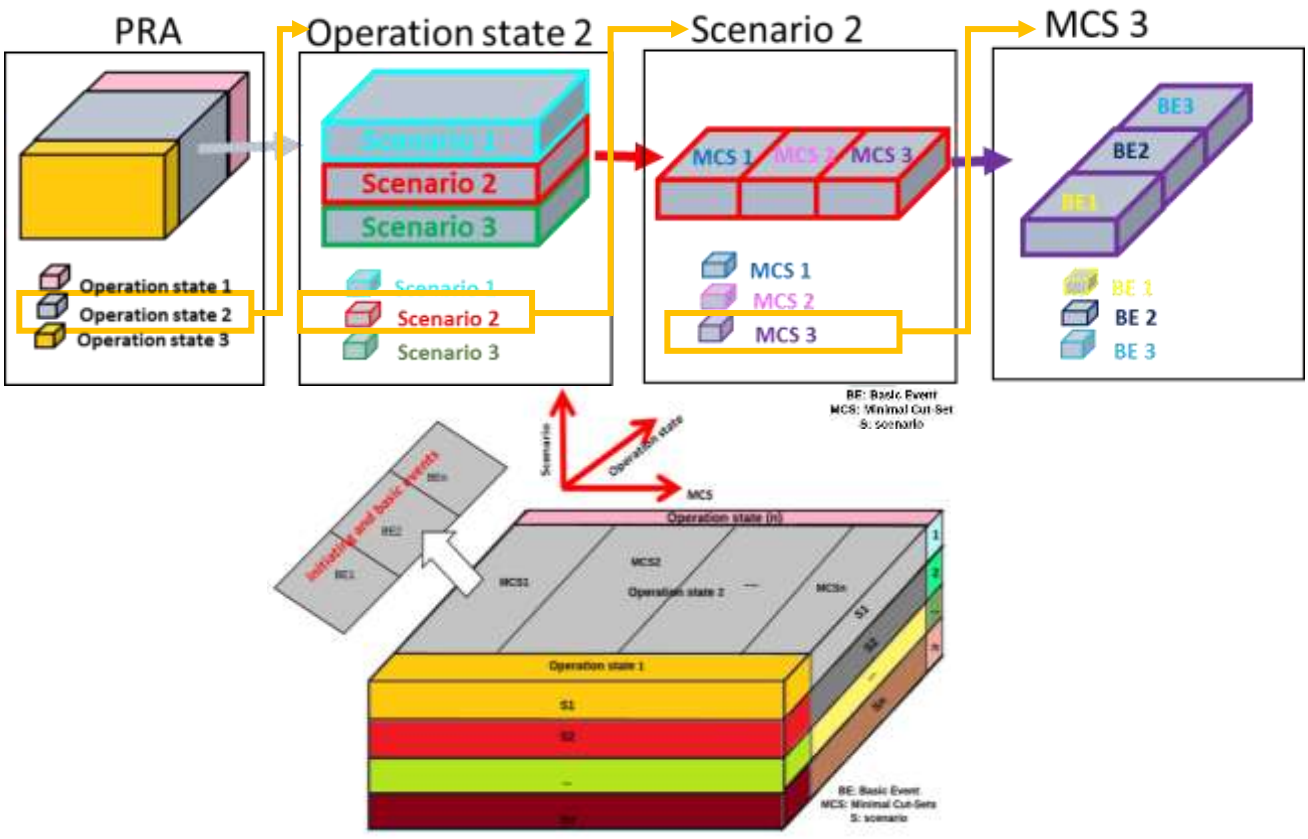


Figure 3 Atomic elements of a PRA model

In Figure 3, let us imagine that the PRA model is a box (cuboid). The box is divided into several cuboids, each representing a given operation state. Each operation state cuboid is further broken down into smaller cuboids that represent the scenarios. The scenario cuboids are in turn broken into smaller cuboids, each representing a MCS. Finally, the MCS cuboids are broken down into the smallest constituting cuboids (known as the basic atomic elements) that represent the basic events. The idea behind this is to facilitate the process of SoK evaluation by decomposing the PRA model into the smallest constituting elements, here called the atomic elements. As illustrated in Figure 3, the atomic elements of the PRA model are the basic events.

To assess the SoK of the PRA model, all the four atomic elements must be considered. In practice, however, PRA models are very complex: they contain many scenarios and operation states, combined in large and complex fault trees and event trees, that consist of thousands of BEs and MCSs (RELCON AB 2005). For such complex risk assessment models, it is not practical to consider all atomic elements for evaluating the SoK. To address this problem, we develop a top-down bottom-up method for SoK assessment, as shown in Figure 4. A reduced-order model for Eq. (1) is developed first, in order to limit the number of atomic elements that need to be analyzed. The model allows the assessment of SoK for most basic atomic elements and, then, calculating it for the other constituting elements. A detailed discussion on how to construct the reduced-order model is given in Section 3.2. Then, the SoK supporting each atomic element in the reduced-order model is assessed by a weighted average of the scores for the attributes in Figure 2. The weights are evaluated using the pairwise comparison matrices of the Analytical Hierarchy Process (AHP), as illustrated in Section 3.3. Finally, the SoK of each element is aggregated to evaluate the SoK of the entire PRA model, which is discussed in details in Section 3.4.

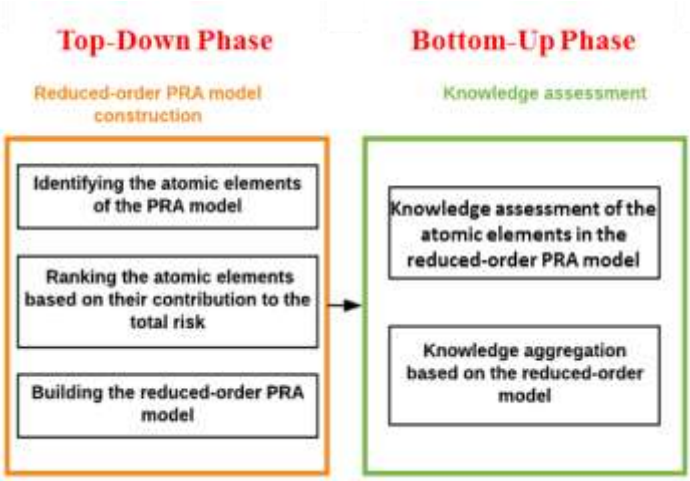


Figure 4 Procedural steps of the developed method

3.2 Reduced-order PRA model construction

In PRA models, most of the contribution to the total risk is provided by a small number of basic elements (known as "*Pareto principle*") (Koch 2011). The rest of the basic elements might be in large number but contribute little to the total risk. To make feasible the SoK assessment, the PRA model is transformed into a reduced-order model that consists of the most important "*atomic elements*", in order to reduce the number of elements that need to be analyzed.

The procedure for constructing the reduced-order model is made of three steps. Firstly, the number of operation states n_O is reduced to the $n_{O,Red}$ most relevant; to do this:

- Calculate the risk R_{O_i} for each operation state:

$$R_{O_i} = \sum_{j=1}^{n_{s,i}} R_{i,j}\,, \quad 1 \le i \le n_O, \tag{3}$$

where $R_{i,j}$ is calculated by (2).

- Rank R_{O_i} $1 \le i \le n_O$ in descending order.
- Find the minimal $n_{O,Red}$, so that:

$$\frac{\sum_{i=1}^{n_{O,Red}} R_{O_i}}{R} \ge \alpha, \tag{4}$$

where α is the fraction of total risk that is represented by the operation states kept in the reduced-order model (in the case study in Section 4, we choose $\alpha = 0.8$).

- Keep only the first, most contributing operation states, i.e., those with $i = 1, \cdots, n_{O,Red}$; operation states with $i > n_{O,Red}$ are eliminated.

The second step is to define the reduced number of scenarios $n_{S,Red,i}$ for each operating state i in the reduced-order model, where $i = 1, \cdots, n_{O,Red}$:

- Calculate the risk $R_{i,j}$, $1 \le j \le n_{S,i}$ by (2).
- Rank $R_{i,j}$ in descending order, $1 \le j \le n_{S,i}$.
- Find the minimal $n_{S,Red,i}$ so that:

$$\frac{\sum_{j=1}^{n_{S,Red,i}} R_{i,j}}{R_{O,i}} \ge \beta, \tag{5}$$

where R_{O_i} is calculated by (3) and β is the fraction of total risk provided by the scenarios in the reduced-order model (in the case study in Section 4, we choose $\beta = 0.8$).

- Keep only scenarios for $j = 1, \cdots, n_{S,Red,i}$; scenarios with $j > n_{S,Red,i}$ are eliminated.
- Repeat the procedures for $i = 1,2, \ldots, n_{O,Red}$.

Finally, the number of minimal cut sets $n_{MCS,i,j}$ is tailored to $n_{MCS,Red,i,j}$, $i = 1, \cdots, n_{O,Red}, j = 1, \cdots, n_{S,Red,i}$:

- Calculate $R_{i,j,k}$ by:

$$R_{i,j,k} = \prod_{q \in MCS_{i,j,k}} P_{BE,q}, \begin{matrix} 1 \le i \le n_{O,Red} \\ 1 \le j \le n_{S,Red,i} \\ 1 \le k \le n_{MCS,i,j} \end{matrix}, \tag{6}$$

- Rank $R_{i,j,k}$ in descending order.
- Find the minimal $n_{MCS,Red,i,j}$ so that:

$$\frac{\sum_{k=1}^{n_{MCS,Red,i,j}} R_{i,j,k}}{R_{i,j}} \ge \gamma, \tag{7}$$

where $R_{i,j,k}$ is calculated by (6) and γ is the fraction of total risk given by the minimal cutsets contained in the reduced-order model (in the case study in Section 4, we choose $\gamma = 0.8$).

- Keep only minimal cut sets for $k = 1, \cdots, n_{MCS,Red,i,j}$; minimal cut sets with $k > n_{MCS,Red,i,j}$ are eliminated.

Taking the rare-event approximation, the total risk of the reduced-order PRA model can be calculated by:

$$R_{Red} = \sum_{i=1}^{n_{O,Red}} \sum_{j=1}^{n_{S,Red,i}} \sum_{k=1}^{n_{MCS,Red,i,j}} \prod_{q \in MCS_{i,j,k}} P_{BE,q}, \quad (8)$$

Only the events that are contained in the reduced-order model (9) are considered when assessing the SoK. Note that from (4), (5) and (7), the reduced order risk R_{Red} accounts for a portion $\alpha \times \beta \times \gamma$ of the total risk R. From (8), the risk index of the reduced-order PRA model can be viewed as the sum of $n_l = \sum_{i=1}^{n_{O,Red}} n_{S,Red,i}$ risk index values $R_{Red,l}, l = 1, \cdots, n_l$ where $R_{Red,l}$ is known as the "elementary risk model" and calculated by the corresponding individual risk model, composed of MCSs and BEs at a given operation state and a given scenario, as shown in (9):

$$R_{Red,l} = \sum_{k=1}^{n_{MCS,Red,l}} \prod_{q \in MCS_{l,,k}} P_{BE,q}, \quad (9)$$

In (9), $R_{Red,l}$ is the risk index of the l-th "elementary reduced-order risk model", where $n_{MCS,Red,l}$ is the number of MCSs in the l-th individual reduced-order risk model. In other words, the "individual reduced-order risk model" represents the risk model at a given operation state and a given scenario.

3.3 SoK assessment for the basic events

The assessment of SoK starts from determining the SoK for each basic event. The total SoK for the reduced PRA model is evaluated as a weighted average of the BEs' SoK, as will be illustrated later in section 3.4. As illustrated previously, the SoK is evaluated as a weighted average of the attributes scores presented in Figure 2, where the attribute scores are evaluated based on the scoring guidelines presented in the Appendixes:

$$K = \sum_{i=1}^{n_i} \sum_{j=1}^{n_{ij}} \sum_{k=1}^{n_{ijk}} W_i . W_{ij} . W_{ijk} . K_{ijk}, \quad (10)$$

In Eq. (10), W_i, W_{ij} and W_{ijk} are respectively the weights of the 2nd, 3rd and 4th level attributes in the hierarchical tree of Figure 2, K_{ijk} is the score of the "leaf" attributes, while n_i, n_{ij} and n_{ijk} are respectively the number of attributes in the 2nd, 3rd and 4th levels. Letting $K_{leaf,k}$ denote the knowledge score for the i-th leaf attribute in the bottom level, Eq. (10) can be simplified as:

$$K = \sum_{k=1}^{n_{leaf}} W_{global,k} . K_{leaf,k}, \quad (11)$$

where $n_{leaf} = 19$ is the number of leaf attributes in the assessment framework of Figure 2, $K_{leaf,k}$ is evaluated based on the guidelines in Appendices A-C, $W_{global,k}$ is the global weight of the k-th "leaf" attribute with respect to the top level goal and is calculated by:

$$W_{global,k} = \begin{cases} W_i . W_{ij}, & if K_{leaf,k} \text{ is in level 3} \\ W_i . W_{ij} . W_{ijk}, & if K_{leaf,k} \text{ is in level 4'} \end{cases} \quad (12)$$

Note that the global weights $W_{global,k}, k = 1,2, \ldots, n_{leaf}$ of the leaf attributes sums to one: $\sum_{k=1}^{n_{leaf}} W_{global,k} = 1$.

As shown in Appendices A-C, $K_{leaf,k}$ is between 1 and 5, with a high value indicating strong knowledge. From Eqs. (10) and (11), it is obvious that also $K_{BE} \in [1,5]$ and a large value indicates strong knowledge on the corresponding BE.

Given the assessment framework developed in Figure 2, the AHP (Saaty 2008) is adopted for evaluating the relative importance (weights) W_i, W_{ij} and W_{ijk} in Eq. (12), due to its capability of considering both quantitative and qualitative evaluations of attributes and factors (Alexander 2012), (Saaty 2008). The AHP method is used for decreasing the complexity of the comparison process for decision-making purposes, as it allows comparing only two criteria (or alternatives) at a time and, then, computing the "overall" relative importance of a criterion in a group of criteria. In addition, it allows gauging and enhancing the rationality and consistency of the expert's evaluation for the criteria, by measuring the consistency of the pairwise comparison matrices. Then, the local relative importance of different alternatives are compared with respect to given criteria and finally, the decision is made based on the overall relative importance of each alternative (Mu & Pereyra-Rojas 2017). However, since there are no alternatives to be compared in this work, pairwise comparison matrices are only needed for deriving the criteria (attributes) weights.

Pairwise comparisons are performed to determine the relative importance (weights) of different attributes (criteria) by comparing their contributions in defining their "parent" attribute (Saaty & Vargas 2012), (Saaty 2008), (Zio 1996). In the application of the method to the case study of the following Section 4, three experts were invited to fill pairwise comparison matrixes. The evaluation scale of Saaty (2008) and Zio (1996) was slightly modified, and a scale of 1-5 was chosen to compare the importance of the attributes with each other. In this scale, two alternatives A and B are compared as the following:

1: A score of (1) is given if A and B are equally important,

2: A score of (2) is given if A is slightly more important than B,

3: A score of (3) is given if A is moderately more important than B,

4: A score of (4) is given if A is strongly more important than B,

5: A score of (5) is given if A is extremely more important than B.

Each expert is asked to fill individually the pairwise comparison matrices, as illustrated above. For each given matrix, the weight of each attribute can, then, be determined by solving the eigenvector problem and normalizing the principal eigenvectors (for details, see (Saaty 2008), (Saaty & Vargas 2012), (Mu & Pereyra-Rojas 2017)). A good approximation to multiply the elements in each row and, then, the n-th root of this product (n is the matrix size) is taken to represent the weight. The output of the row is eventually, normalized with the other row's outputs. For more details on AHP and deriving the weights from pairwise comparison matrices, see: (Coyle 2004), (Saaty 2013).

It should be noted that the consistency of the pairwise comparison matrix should be checked by calculating the consistency ratio (CR):

$$CR = \frac{CI}{RI}, \tag{13}$$

where RI represents the consistency index of a randomly generated matrix and its value can be taken from Table 1 of Saaty and Tran (2007), and CI is the consistency index which is calculated by (14):

$$CI = \frac{\lambda_{max} - n}{n-1}, \tag{14}$$

where λ_{max} is the maximum eigenvalue and n is the order of the matrix and represents the number of attributes being compared (Saaty 2008), (Zio 1996). Saaty's acceptance criteria of consistency is adopted (Saaty 2008): when $CR < 0.1$, the comparison matrix is consistent, otherwise it is not and the experts are demanded to revise their evaluations (Zio 1996) (Alonso & Lamata 2006), (Saaty & Tran 2007). After checking the consistency of the matrices and obtaining the weights of the attributes from each expert, the final weight of each attribute is calculated by averaging the weights obtained from the experts.

As illustrated in Sect 3.2, the PRA model is deconstructed to its constituting elements and then, the number of constituting elements is reduced. In this reduced order PRA model, the most basic element is the "basic event", where a minimal cutset consists of a group of "basic events". On the other hand, a given scenario mathematically consists of a group of minimal cutsets. Finally, a given operation states consist of a group of scenarios. Accordingly, the assessment of the SoK starts with the evaluation of the BEs in the reduced-order model of Eq. (8). The SoK of the BEs is denoted by K_{BE} and evaluated as in Eq. (11) by a weighted average of the leaf attributes scores. We take the generic q-th BE as an example to illustrate step by step the evaluation of the SoK assessment method. For the sake of simplicity, we dropped the q subscripts in the symbols:

$$K_{BE} = \sum_{k=1}^{n_{leaf}} W_{global,k} . K_{leaf,k} \tag{15}$$

3.4 Aggregation of the SoK

Once the SoKs of the basic events in the reduced-order models are evaluated, they can be aggregated to evaluate the total SoK for the PRA model. Let $K_{BE,l,q}$ represent the SoK of the q-th BE in the l-th reduced-order model. The aggregation of $K_{BE,l,q}$ should consider the difference in the atomic elements' (i.e., BEs, MCs, Scenarios, etc.) contribution to the total risk. Different importance measures can be used to evaluate the contribution of the basic events. For example, as the reduced-order risk model is constructed by the BEs in the MCSs, the weights of the BEs can be calculated based on Fussell-Vesely importance measures (Zio 2007):

$$W_{BE,l,q} = \frac{I_{BE,l,q}}{\sum_{q=1}^{n_{BE,l}} I_{BE,l,q}}, \tag{16}$$

where $I_{BE,l,q}$ is the Fussell-Vesely importance measure value of the corresponding q-th BE in the elementary risk model l. Remember that the "elementary reduced-order risk model" represents the risk model at a given operation state and a given scenario, and it is composed of the sum of MCSs (computed by the BEs) in this scenario, as illustrated in Eq.(9).

The SoK for the l-th elementary reduced-order risk model, denoted by K_l, is calculated by a weighted average of knowledge scores on its basic events by:

$$K_l = \sum_{q=1}^{n_{BE,l}} W_{BE,l,q} \cdot K_{BE,l,q}, \tag{17}$$

The importance of the reduced-order model is evaluated by its contribution to the total risk:

$$W_l = \frac{R_{Red,l}}{\sum_{l=1}^{n_l} R_{Red,l}}, \tag{18}$$

where $R_{Red,l}$ is the risk index value of the l-th "elementary reduced-order model" and is calculated by (9).

To calculate the total SoK K_{Red} of the reduced-order risk model, the knowledge indexes K_ls of the individual reduced-order risk models are further aggregated by considering their contributions:

$$K_{Red} = \sum_{l=1}^{n_l} W_l K_l, \tag{19}$$

The index K_{Red} is, then, used to represent the SoK of the entire PRA of a specific hazard group: its value is between 1 and 5, with a high value indicating that there is strong knowledge in support of the PRA model and its risk outcomes.

4. Case study

In this section, we apply the developed framework to a case study of real PRA models for two hazard groups in NPPs. The reduced-order model is constructed first for each hazard group. The SoK assessment framework is, then, applied on the BEs and the total SoK is obtained by aggregating the BEs' SoKs. Finally, a comparison is made on the SoKs of the two PRA models to provide some conclusions to relevant RIDM.

4.1 Description of PRA models

In this section, we consider a case study extracted from PRA models of two hazard groups, i.e., external flooding and internal events provided by Electricité De France (EDF). Both PRA models were developed using the Risk Spectrum Professional software.

In all generality, "external hazards" refer to undesired events originating from sources outside the NPP, such as external flooding, external fires, seismic hazards etc. (IAEA 2010). In this paper, we consider a particular external hazard, i.e., external flooding, that is caused by the overflow of water due to naturally induced external causes, e.g., tides, tsunamis, dam failures, snow melts, storm surges, etc. (IAEA 2003). The "external flooding" PRA model considered in this paper is a combination of event trees and fault trees that are constructed to evaluate the risk of external flooding in different water level conditions (scenarios). The total risk index of external flooding is, then, calculated by summing the risk indexes at each water level. The PRA model of external flooding is complex and has a large scale, including three operation states, thousands of BEs and several thousands of MCSs.

"Internal events" refer to undesired events that originate within the NPP itself and can cause initiating events that might lead to loss of important systems and, eventually, a core meltdown (EPRI 2015). Major internal events include componenets, systems or structural failures, safety systems operation, and maintenance errors, etc. (IAEA Safety Standards Series 2009). Internal events might also lead to other initiating events like turbine trip and Loss of Coolant Accidents (LOCAs). In nuclear PRA, internal events are considered a well-established and understood hazard group (EPRI 2012), and highly mature PRA models are available for their characterization. The internal events PRA model considered in this paper is based on a combination of event trees and fault trees that are constructed for evaluating the risk over different internal events (e.g., loss of offsite power, loss of auxiliary systems). The risk index of the entire internal events hazard group is, then, calculated by summing the risk indexes (i.e., minimal cut sets at a given operation state and scenario) of the individual internal events. Similarly to the PRA model of external flooding, the PRA model of internal events is complex and has a large scale, also containing three operation states, few thousands of BEs and several thousands of MCSs.

4.2 Reduced-order model construction

The first step in the developed SoK assessment method is the reduced-order model construction. Here, we only show in details how to construct the reduced-order risk assessment model for the external flooding PRA model. For the internal events PRA model, the reduced-order model can be constructed in a similar way.

In this paper, we set the fractions of the risk to be $\alpha = \beta = \gamma = 0.8$. From Eq. (4), we found that only one out of six operation states (NS/SG-normal shutdown with cooling using steam generator-NS/SG) is needed for the reduced-order model, which contributes to 86% of the total risk index. Therefore, we have $n_O = 1$. Similarly, based on Eq. (5), only one out of ten scenarios (water levels) is needed for the reduced-order model, whose risk contribution is 98.7%. Hence, we have $n_S = 1$. Based on Eq. (7), given the operation states and scenarios of interest, 5 out of 3102 MCSs already contribute to 80.1% of the risk at the given operation state and scenario. Thus, we have $n_{MCS} = 5$. Then, a reduced-order model can be constructed using the atomic elements in Table 4. The definitions of BEs in the MCSs of Table 4 can be found in Table 5. An illustration example on the pathway of the first minimal cut sets is given in Figure 5. Assuming the rare-event approximation, the risk index of interest, i.e., the probability of core meltdown, can be calculated using the MCSs and the BEs in Table 4, following Eqs. (4), (5), (7) and (8). The constructed reduced-order risk model can reconstruct $86\% \times 98.7\% \times 80.1\% = 67.99\%$ of the total risk R.

Table 4 Reduced-order model constituents

Operating state	Scenarios	MCS
NS/SG	Water level A	MCS1={BE1, BE2, BE3} MCS2={BE2, BE3, BE4} MCS3={BE3, BE5, BE6, BE7, BE8} MCS4={BE2, BE3, BE7, BE9}

	MCS5={ BE2, BE3, BE6, BE10}

Table 5 Basic events included in the reduced-order model

Symbol	Basic event
BE1	External flooding with water level A inducing a loss of offsite power
BE2	Loss of auxiliary feedwater system due to the failure to close the isolating valve
BE3	Loss of component cooling system because of clogging
BE4	Failure of all pumps of the Auxiliary feedwater (AFW) system
BE5	Failure of the turbine of the AFW system
BE6	Failure of the Diesel Generator A
BE7	Failure of the Diesel Generator B
BE8	Failure of the common diesel generator
BE9	Failure of pumps 1 and 2 of AFW system
BE10	Failure of pumps 2 and 3 of AFW system

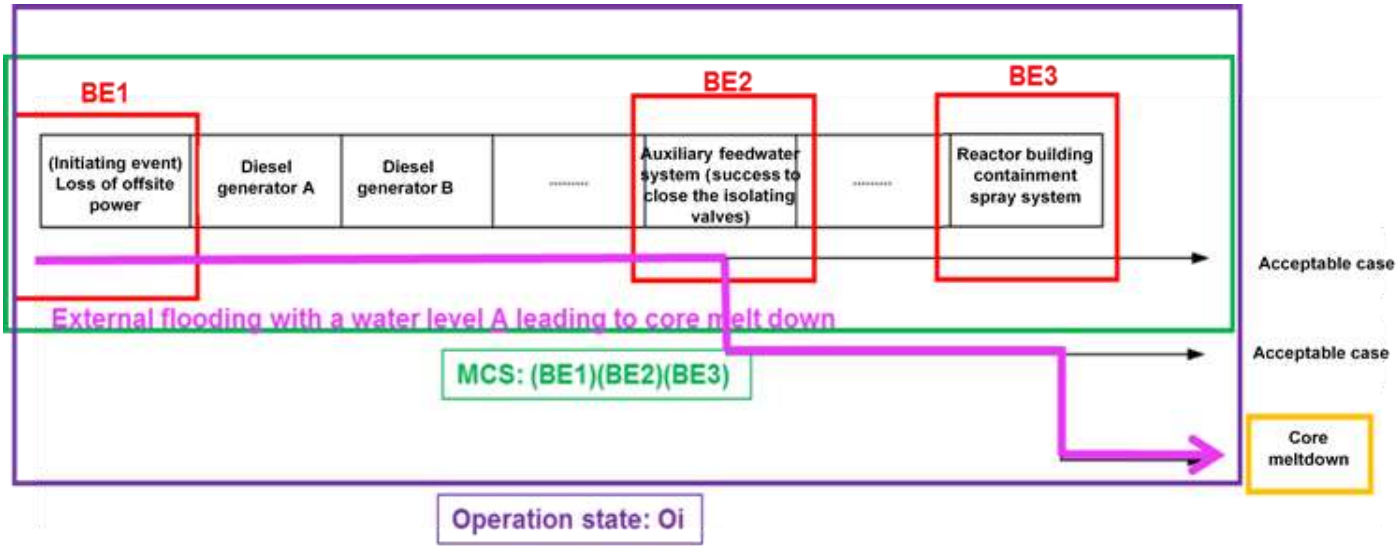


Figure 5 Illustration of a MCS in an individual reduced-order model

4.3 Knowledge assessment of basic events

In this section, we show how to assess the SoK for the BEs in Table 5. As shown in Eq. (11), the SoK of the basic event is evaluated as a weighted average over the SoK of the 19 leaf attributes in Figure 2. Hence, the first step of applying the SoK assessment framework is to determine the global weights of the "leaf" attributes. It should be noted that these weights are the same for all basic events. Hence, this step needs to be done only once. Take the "leaf" attribute K_{31} (years of experience) as an example. From Figure 2, it can be seen that K_{31} shares the same parent with the other three attributes K_{32}, K_{33} and K_{34}. To identify its global weight, a 4×4 pairwise matrix needs to be constructed by experts to compare the importance of the three attributes with respect to their parent attribute. The results of the pairwise comparison matrix is given in Table 6. In this matrix, the score $S_{1,2} = 3$ in the first raw, means that YE is more important that NE.

Table 6 Pairwise comparison matrix for the assumptions daughter attributes of K_1 (expert 1)

A	YE	NE	AE	In	W
YE	1	4	1	1	0.318
NE	1/3	1	1/3	1/3	0.092
AE	1	3	1	1	0.295
In	1	3	1	1	0.295

After constructing the pairwise comparison matrix, the consistency of the matrix needs to be checked. The maximum eigenvalue of the matrix is $\lambda_{max} = 4.082$; the consistency index for the matrix ($n = 4$) is, then, calculated according to Eq. (14) to be $CI = 0.027$. From Table 1 in Saaty and Tran (2007), the random

index is RI =0.89. The consistency ratio is, then, found by Eq. (13) to be $CR = 0.031$: since $CR < 0.1$, the consistency of the matrix is accepted. The weight of each attribute is, then, found by normalizing the principal eigenvector, following the instructions in Section 3.3. The weight of the parent attribute K_3 (understanding of phenomena) was found to be $W_3 = 0.306$. The global weight for K_{31} of the leaf attributes can, be determined using Eq. (12): $K_{31} = W_3.W_{31} = 0.097$. The experts were asked to repeat the same steps. The weights obtained for each leaf attribute from each expert were then averaged. The results are presented in Tables 7-8.

Then, the SoK for the "leaf" attributes, i.e., $K_{leaf,i}$ in Eq. (11) is determined following the assessment guidelines in Appendices A-C. Here, we give an illustrating example on how to evaluate the SoK of the basic event BE_2. The first leaf attribute, i.e., quality of assumptions K_{11}, is evaluated based on the guidelines in Appendix A.1. In this basic event, the loss of equipment is calculated by assuming that as long as the water reaches the bottom of each equipment, a failure is caused. This assumption is based on extrapolating some data to extreme values, and it is conservative. Therefore, this assumption was judged by the experts to lie between two cases with score 1 and score 3 in Table A.1: an inter-level score of 2 was given by the experts. Take the amount of data K_{21} as another example: the number of years of experience on BE_2 is 10 years; therefore, from Appendix B.1, the SoK score of K_{21} is assessed by the experts to be 1. The rest of the leaf attributes are assessed similarly and the results are given in Table 7 and Table 8. Then, from Eq. (11) we found $K_{BE} = 3.5500$ for BE_2. The procedures are repeated for each BE; the resulting K_{BE}s are given in Table 9.

Table 7 Assessment of level-3 knowledge "leaf" attributes (BE_2)

Attribute	QA	AD	YE	NE	AE	IN
$W_{i,global}$	0.3234	0.0587	0.1190	0.0630	0.1190	0.1190
Score	2	1	5	5	5	5

Table 8 Assessment of level-4 knowledge "leaf" attributes (BE_2)

Attribute	PK	SI	U	RI	PE	PM	P	C	Co	V	Cu	Ac
$W_{global,k}$	0.0203	0.0134	0.0177	0.0144	0.0179	0.0186	0.0221	0.0148	0.0110	0.0147	0.0139	0.0190
Score	5	5	4	4	5	5	4	5	5	3	4	3

4.4 Knowledge Aggregation

Finally, the K_{BE}s in Table 9 are aggregated for the SoK of the entire model. For this, the SoK of the individual reduced-order risk models K_l need to be calculated first by Eqs. (16) and (17), with the Fussell-Vesely (FV) importance measures for the BEs also given in Table 9. In this case study, we have $l = 1$ for the external events. The resulted K_l from Eqs. (16) and (17) is $K_l = 2.90$. Then, the total SoK for external flooding, denoted by $K_{Red,Ex}$, is calculated based on the reduced-order model using Eqs. (18) and (19). In

this case study, since we have only one individual risk model, using Eqs. (18) and (19) leads to $K_{Red,Ex} = K_{l,1} = 2.90$.

Table 9 Knowledge assessment and aggregation over the basic events

BE	BE1	BE2	BE3	BE4	BE5	BE6	BE7	BE8	BE9	BE10
FV	0.9020	1.0000	0.5530	0.1820	0.1410	0.1270	0.1210	0.0450	0.0277	0.0277
$W_{BE,l,q} = NFV$	0.2885	0.3199	0.1769	0.0582	0.0451	0.0406	0.0387	0.0144	0.0089	0.0089
K_{BE}	1.6582	3.6595	2.9006	3.2178	3.7778	3.7778	3.0102	3.7778	3.2178	3.2178
$W_{BE,l,q} \times K_{BE,l,q}$	0.4784	1.1705	0.5131	0.1873	0.1704	0.1535	0.1165	0.05437	0.0285	0.0285

*(FV): Fussell-Vesely

*(NFV): Normalized Fussell-Vesely

4.5 Results and discussion

The same steps were repeated on the internal events PRA model. We directly present the final SoK for the internal events PRA model: $K_{Red,In} = 4.04$. The SoK for both hazard groups are graphically illustrated in Figure 6. In Figure 6, we also illustrate the risk indexes (probability of core meltdown) evaluated for the two hazard groups (note that the values of the risk indexes are scaled due to confidentiality reasons). It can be seen from the Figure 6 that the SoK on the internal events is higher than that on external flooding: this means that we are surer of the risk index value calculated with the PRA model of internal events, than of that for the external flooding hazard group.

In fact, these results confirm expectations, as the internal events hazard group has been well studied in nuclear PRAs and mature models are available, whose parameters have relatively low uncertainty (EPRI 2015). On the other hand, the PRAs for external flooding is generally considered less mature (EPRI 2012) and several limitations have been pointed out in the current external flooding PRA models. For example, the flood frequencies are obtained by extrapolating the fitted historical data (usually limited) to the design basis flood levels, which results in high uncertainty (EPRI 2012). In particular, the probability of extreme floods is very low (IAEA 2003) and flooding events are very site-specific (IAEA 2009). Hence, very few data are available for risk modeling, which limits the SoK for external flooding. The low occurrence probability of external flooding and the lack of operating experience and data related to them makes it very difficult also to predict and estimate their consequences, which adds to the uncertainties in the risk analysis as it limits the SoK of the PRA model used (IAEA 2003). Specifically, in the case study considered, a large fraction of the risk contribution (69% of the reduced-order risk for external flooding) is due to three basic events i.e., BE_1, BE_2, and BE_3. As shown in Table 9, two of them (BE_1, BE_3) have quite low SoK, which limits the SoK of the entire PRA model.

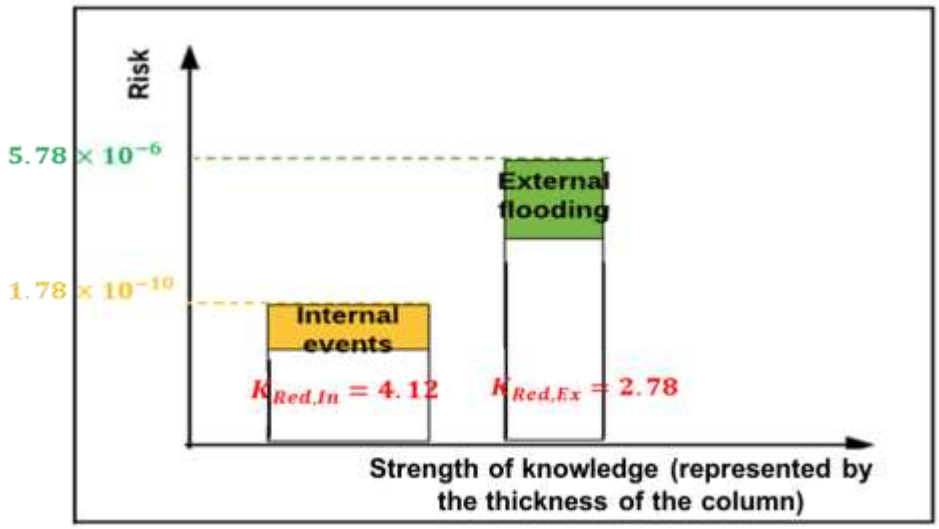


Figure 6 Representation of hazard groups levels of risk and SoK

1. Conclusions

In this paper, we have proposed a new method for implementing a quantitative evaluation of the SoK of risk assessment models. The underlying conceptual framework has been developed based on a thorough literature review. The framework is based on three main attributes (assumptions, data, and phenomenological understanding), which are further decomposed into more tangible sub-attributes and "leaf" attributes for quantification. Detailed scoring guidelines are defined for the evaluation of the leaf attributes. In order to facilitate the application of the knowledge evaluation framework in practice, a top-down bottom-up approach is proposed, where a reduced-order model is constructed in the top-down phase to reduce the complexity of the analysis, and the SoKs are evaluated and aggregated hierarchically in the bottom-up phase. The application of the framework on a real case study of PRA models for two hazard groups, i.e., external flooding and internal events in NPP, has shown its operability. The results of the case study are consistent with the expectations of industrial practice, where the SoK of external flooding is lower than that of internal events, for which more data and information (i.e., strong knowledge) are available.

A potential limitation of the developed method is that we are assuming that the risk assessment model itself is complete in covering all the possible scenarios. The SoK on model structure and model uncertainty (Droguett & Mosleh 2008), (Droguett 1999) is not considered in this paper. For a more comprehensive knowledge assessment, further studies are needed to extend the developed method to consider completeness and comprehensiveness, including model uncertainty in the PRA model (Droguett & Mosleh 2008), (Droguett 1999). Also, as the weights of the attributes in the framework are subjectively evaluated, formal expert judgment elicitation methods should be used for evaluating the weights. Finally, the evaluation framework and method do not pretend to be complete but they stand as a starting point for a practical assessment of the SoK of risk assessment models.

References

1. Abrahamsen, H.B., Abrahamsen, E.B. & Høyland, S., 2016. On the need for revising healthcare failure

mode and effect analysis for assessing potential for patient harm in healthcare processes. *Reliability Engineering & System Safety*, 155, pp.160–168. Available at: http://www.sciencedirect.com/science/article/pii/S095183201630179X.

2. Alexander, M., 2012. Decision-Making using the Analytic Hierarchy Process (AHP) and SAS/ IML. *The United States Social Security Administration Baltimore*, pp.1–12.
3. Alonso, J.A. & Lamata, M.T., 2006. Consistency in the analytic hierarchy process: a new approach. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 14(4), pp.445–459. Available at: http://www.worldscientific.com/doi/abs/10.1142/S0218488506004114.
4. Apostolakis, G., 1990. The concept of probability in safety assessments of technological systems. *Science*, 250(4986), pp.1359–1364.
5. Askeland, T., Flage, R. & Aven, T., 2017. Moving beyond probabilities ??? Strength of knowledge characterisations applied to security. *Reliability Engineering and System Safety*, 159(October 2016), pp.196–205.
6. Aven, T., 2013a. A conceptual framework for linking risk and the elements of the data-information-knowledge-wisdom (DIKW) hierarchy. *Reliability Engineering and System Safety*, 111, pp.30–36.
7. Aven, T., 2017a. How some types of risk assessments can support resilience analysis and management. *Reliability Engineering & System Safety*, 167, pp.536–543.
8. Aven, T., 2017b. Improving risk characterisations in practical situations by highlighting knowledge aspects, with applications to risk matrices. *Reliability Engineering & System Safety*, 167, pp.42–48. Available at: http://www.sciencedirect.com/science/article/pii/S0951832016306950.
9. Aven, T., 2013b. Practical implications of the new risk perspectives. *Reliability Engineering and System Safety*, 115, pp.136–145.
10. Aven, T., 2012. The risk concept historical and recent development trends. *Reliability Engineering & System Safety*, 99, pp.33–44. Available at: http://www.sciencedirect.com/science/article/pii/S0951832011002584.
11. Aven, T. & Krohn, B.S., 2014. A new perspective on how to understand, assess and manage risk and the unforeseen. *Reliability Engineering & System Safety*, 121, pp.1–10. Available at: http://www.sciencedirect.com/science/article/pii/S0951832013002159.
12. Aven, T. & Ylönen, M., 2016. Safety regulations: Implications of the new risk perspectives. *Reliability Engineering & System Safety*, 149, pp.164–171. Available at: http://www.sciencedirect.com/science/article/pii/S0951832016000168.
13. Bani-Mustafa, T. et al., 2017. A hierarchical tree-based decision making approach for assessing the trustworthiness of risk assessment models. In *PSA (ANS)*. American Nuclear Society (ANS).
14. Bergdahl, M. et al., 2007. *Handbook on Data Quality Assessment Methods and Tools*,
15. Berner, C. & Flage, R., 2016a. Comparing and integrating the NUSAP notational scheme with an uncertainty based risk perspective. *Reliability Engineering & System Safety*, 156, pp.185–194.
16. Berner, C. & Flage, R., 2016b. Strengthening quantitative risk assessments by systematic treatment of uncertain assumptions. *Reliability Engineering and System Safety*, 151, pp.46–59.
17. Bjerga, T. & Aven, T., 2015. Adaptive risk management using new risk perspectives – an example from the oil and gas industry. *Reliability Engineering & System Safety*, 134, pp.75–82. Available at:

http://www.sciencedirect.com/science/article/pii/S0951832014002531.

18. Boone, I. et al., 2010. NUSAP: a method to evaluate the quality of assumptions in quantitative microbial risk assessment. *Journal of Risk Research*, 13(3), pp.337–352. Available at: http://www.scopus.com/inward/record.url?eid=2-s2.0-77951165131&partnerID=40&md5=12a3caae6ff5f3fae9967becb6b35f17.
19. Chen, H. et al., 2014. A review of data quality assessment methods for public health information systems. *International Journal of Environmental Research and Public Health*, 11(5), pp.5170–5207.
20. Coyle, G., 2004. The analytic hierarchy process (AHP). *Practical strategy: Structured tools and techniques*, pp.1–11.
21. DAMA, 2013. *The six primary dimensions for data quality assessment: defining data quality dimensions*,
22. Davies, M., 2001. Knowledge (explicit and implicit): philosophical aspects.
23. Droguett, E.L., 1999. *Methodology for the treatment of model uncertainty*,
24. Droguett, E.L. & Mosleh, A., 2008. Bayesian methodology for model uncertainty using model performance data. *Risk Analysis*, 28(5), pp.1457–1476.
25. EPRI, 2015. *An Approach to Risk Aggregation for RisK-Informed Decision-Making*, Palo Alto, California.
26. EPRI, 2012. *Practical Guidance on the Use of Probabilistic Risk Assessment in Risk-Informed Applications with a Focus on the treatment of Uncertanity*, Palo Alto, California.
27. Flage, R. & Aven, T., 2009. Expressing and communicating uncertainty in relation to quantitative risk analysis. *Reliability: Theory & Applications*, 4(2–1 (13)).
28. Goerlandt, F. & Montewka, J., 2014. Expressing and communicating uncertainty and bias in relation to Quantitative Risk Analysis. *Safety and Reliability: Methodology and Applications*, 2(13), pp.1691–1699. Available at: http://www.crcnetbase.com/doi/abs/10.1201/b17399-230.
29. Goerlandt, F. & Montewka, J., 2015. Maritime transportation risk analysis: Review and analysis in light of some foundational issues. *Reliability Engineering & System Safety*, 138, pp.115–134. Available at: http://www.sciencedirect.com/science/article/pii/S0951832015000356.
30. Goerlandt, F. & Reniers, G., 2016. On the assessment of uncertainty in risk diagrams. *Safety Science*, 84, pp.67–77. Available at: http://www.sciencedirect.com/science/article/pii/S0925753515003215.
31. Haouzi, H. El et al., 2013. Toward Adaptive Modelling & Simulation for IMS : The Adaptive Capability Maturity Model and Future Challenges To cite this version :
32. Helton, J.C. & Burmaster, D.E., 1996. Guest editorial: treatment of aleatory and epistemic uncertainty in performance assessments for complex systems. *Reliability Engineering & System Safety*, 54(2–3), pp.91–94.
33. Helton, J.C., Johnson, J.D. & Oberkampf, W.L., 2004. An exploration of alternative approaches to the representation of uncertainty in model predictions. *Reliability Engineering & System Safety*, 85(1–3), pp.39–71.
34. Hey, J., 2004. *The data, information, knowledge, wisdom chain: the metaphorical link. Intergovernmental Oceanographic Commission.*, Available at: http://www.dataschemata.com/uploads/7/4/8/7/7487334/dikwchain.pdf.

35. IAEA, 1991. *Data Collection and Record Keeping for the Management of Nuclear Power Plant Ageing* IAEA, ed.,
36. IAEA, 2010. *Development and Application of Level 1 Probabilistic Safety Assessment for Nuclear Power Plants*,
37. IAEA, 2003. *External Events Excluding Earthquakes in the Design of Nuclear Power Plants*,
38. IAEA, 2009. *Meteorological and Hydrological Hazards in Site Evaluation for Nuclear Installations*,
39. IAEA Safety Standards Series, 2009. *Deterministic Safety Analysis for Nuclear Power Plants*,
40. De Jong, A., Wardekker, J.A. & Van der Sluijs, J.P., 2012. Assumptions in quantitative analyses of health risks of overhead power lines. *Environmental science & policy*, 16, pp.114–121.
41. Kelp, C., 2015. Understanding phenomena. *Synthese*, 192(12), pp.3799–3816.
42. Khorsandi, J. & Aven, T., 2017. Incorporating assumption deviation risk in quantitative risk assessments: A semi-quantitative approach. *Reliability Engineering & System Safety*, 163, pp.22–32.
43. Kidwell, J.J., Vander Linde, K. & Johnson, S.L., 2000. Applying corporate knowledge management practices in higher education. *Educause quarterly*, 23(4), pp.28–33.
44. Kloprogge, P., Van der Sluijs, J.P. & Petersen, A.C., 2011. A method for the analysis of assumptions in model-based environmental assessments. *Environmental Modelling and Software*, 26(3), pp.289–301. Available at: http://dx.doi.org/10.1016/j.envsoft.2009.06.009.
45. Koch, R., 2011. *The 80/20 principle: the secret to achieving more with less*, Crown Business.
46. Milazzo, M.F. & Aven, T., 2012. An extended risk assessment approach for chemical plants applied to a study related to pipe ruptures. *Reliability Engineering & System Safety*, 99, pp.183–192. Available at: http://www.sciencedirect.com/science/article/pii/S095183201100264X.
47. Montewka, J., Goerlandt, F. & Kujala, P., 2014. On a systematic perspective on risk for formal safety assessment (FSA). *Reliability Engineering & System Safety*, 127, pp.77–85. Available at: http://www.sciencedirect.com/science/article/pii/S095183201400057X.
48. Morgan, S.L. & Waring, C.G., 2004. *Guidance on Testing Data Reliability*,
49. Mu, E. & Pereyra-Rojas, M., 2017. Understanding the analytic Hierarchy process. In *Practical Decision Making*. Springer, pp. 7–22.
50. Nowakowski, T. et al., 2014. *Safety and reliability: Methodology and applications*, CRC Press.
51. NRC, U.S., 1983. *PRA procedures Guide*, NUREG/CR-2300.
52. Popek, E.P., 2017. *Sampling and analysis of environmental chemical pollutants: a complete guide*, Elsevier.
53. RELCON AB, 2005. *Theory Manual*,
54. Roth, D.J., 2009. Assessing the Reliability of Computer-Processed Data. , (July).
55. Rowley, J. & Hartley, R., 2017. *Organizing knowledge: an introduction to managing access to information*, Routledge.
56. Saaty, T.L., 2013. Analytic hierarchy process. In *Encyclopedia of operations research and management science*. Springer, pp. 52–64.
57. Saaty, T.L., 2008. Decision making with the analytic hierarchy process. *International Journal of Services Sciences*, 1(1), p.83.
58. Saaty, T.L. & Tran, L.T., 2007. On the invalidity of fuzzifying numerical judgments in the Analytic

Hierarchy Process. *Mathematical and Computer Modelling*, 46(7–8), pp.962–975.

59. Saaty, T.L. & Vargas, L.G., 2012. *Models, methods, concepts & applications of the analytic hierarchy process*, Springer Science & Business Media.
60. Saltelli, A. et al., 2013. What do I make of your latinorum? Sensitivity auditing of mathematical modelling. *International Journal of Foresight and Innovation Policy*, 9(2-3–4), pp.213–234.
61. Van Der Sluijs, J.P. et al., 2005. Combining Quantitative and Qualitative Measures of Uncertainty in Model-Based Environmental Assessment: The NUSAP System. *Risk Analysis*, 25(2), pp.481–492. Available at: http://doi.wiley.com/10.1111/j.1539-6924.2005.00604.x.
62. Stamatelatos, M. et al., 2011. *Probabilistic risk assessment procedures guide for NASA managers and practitioners*,
63. Stirling, P.A., 1999. On Science and Precaution in the Management of Technological Risk: Volume II-case studies.
64. Valdez Banda, O.A. et al., 2015. A risk analysis of winter navigation in Finnish sea areas. *Accident Analysis & Prevention*, 79, pp.100–116. Available at: http://www.sciencedirect.com/science/article/pii/S0001457515000986.
65. Zio, E., 2007. An introduction to the basics of reliability and risk analysis. Series in Quality. *Reliability and Engineering Statistics*, 13.
66. Zio, E., 1996. On the use of the analytic hierarchy process in the aggregation of expert judgments. *Reliability Engineering and System Safety*, 53(2), pp.127–138.

Appendix A: Evaluation guidelines for leaf attributes under Solidity of Assumptions (K_1)

Table A.1 Scoring guidelines for quality of assumptions *(*Boone *et al.,2010)*

Score / Attribute	1	3	5
Quality of assumptions K_{11}	$K_{11} = 1$ if the ssumption is not realistic (over conservative or over optimistic), or the available information is not sufficient for assessing the quality of the assumptions	$K_{11} = 3$ if the assumption is based on existing simple models and extrapolated data	$K_{11} = 5$ if the assumption is plausible: it is grounded on well-established theory or abundant experience on similar systems, and verified by peer review

Note: If multiple assumptions are involved in the assessment, the final score for K_{11} is obtained by averaging the scores of all the assumptions.

Table A.2 Scoring guidelines for the value-ladenness of the assessors

Score / Attribute	1	3	5
Personal knowledge (educational background) K_{121}	$K_{121} = 1$ if all of the experts hold academic degrees from other domains	$K_{121} = 3$ if less than two thirds of the experts hold academic degrees in the same field	$K_{121} = 5$ if over two thirds of the experts hold academic degrees in the same field
Sources of information K_{122}	$K_{122} = 1$ if experts can only access academic information source or only industrial information source	$K_{122} = 3$ if experts can access fully industrial information source and partially academic information source	$K_{122} = 3$ if experts can fully access both academic and industrial information sources
Unbiasedness and plausibility K_{123}	$K_{123} = 1$ if the expert team is very conservative or optimistic	$K_{123} = 3$ if the expert team is slightly conservative/optimistic	$K_{123} = 5$ if as a team, the experts are unbiased: the biases of the experts can compensate one another
Relative independence K_{124}	$K_{124} = 1$ if over three quarters of the experts are highly influenced by mangers and stakeholders	$K_{124} = 3$ if less than one quarter of experts might be influenced by the mangers and stakeholders	$K_{124} = 5$ if all experts' decisions are highly independent

Past experience K_{125}	$K_{125} = 1$ if the experts' experience is less than 5 years	$K_{125} = 3$ if the experts' experience is between 10-15 years	$K_{125} = 5$ if the experts' experience is more than 20 years
Performance measure K_{126}	$K_{126} = 1$ if the performance of the experts are not evaluated by external peers	$K_{126} = 3$ if the external peers generally acknowledge the experts' performance but raise some slight concerns	$K_{126} = 5$ if the external peers endorse the experts' performance and approve them
Agreement among peers K_{127}	$K_{127} = 1$ if some experts hold strongly conflicting views on the assumptions	$K_{127} = 3$ if some experts questions on the assumptions, but do not have strongly conflicting views	$K_{127} = 1$ if most of the experts agree on the assumptions

Table A.3 Scoring guidelines for assumption sensitivity

Score / Attribute	1	3	5
Sensitivity of assumptions K_{13}	$K_{13} = 1$ if the assumption greatly influences the final result	$K_{13} = 3$ if the assumption greatly influences the results in a major step in the calculation	$K_{13} = 5$ if the assumption has little or no impact on the results of risk analysis

Note: The score here is related to the impact of the sensitivity on the SoK

Appendix B: Evaluation guidelines for leaf attributes under Availability and Reliability of Data ($\mathbf{K_2}$)

Amount of data K_{21} is measured by a numerical metric, Years of Experience (YoE), defined by the number of related events recorded during a specific period.

YoE =length of the data collection period (in years) × sample size of the data

The amount of data is scored based on the criteria in Table B.1.

Table B.1 Scoring guidelines for Amount of available data K_{21}

Value of YoE	Score
< 50	1
50-199	2
200-499	3
500-999	4
>1000	5

Completeness of data refers to the degree to which the collected data contains the needed information. For components and systems, data completeness is characterized by the following criteria (IAEA 1991):

1. The data should contain baseline information, which covers the design data and conditions of a component at its initial state.
2. The data should contain the operating history, which covers the service conditions of systems and components including transient and failure data.
3. The data should contain the maintenance history data, which covers the components monitoring and maintenance data.

For more details on how each of the previous attributes is identified, see (IAEA 1991). However, it should be noted that the completeness features are defined differently depending on the problem. For example, data required for quantifying to a component failure frequency is different from that for quantifying a natural event. General scoring guidelines for evaluating K_{221} are given, based on the degree to which criteria are satisfied, as shown in Table B.2.

Table B.2 scoring guidelines for data reliability

Score / Attribute	1	3	5
Completeness K_{221}	$K_{221} = 1$ if the data fail to contain the necessary information required in developing the risk assessment model (in the light of the completeness characteristics defined above)	$K_{221} = 3$ if the data contain to an acceptable degree the necessary information required in developing the risk assessment model (in the light of the completeness characteristics defined above)	$K_{221} = 5$ if the data contain all the necessary information required in developing the risk assessment model (in the light of the completeness characteristics defined above)

The validity of data is evaluated by the following criteria:

1. The integrity of data is carefully managed.
2. Databases are well organized and formatted in a common way, and easily retrieved and manipulated.
3. Data should be collected and entered in the database by well-trained maintenance personnel, and modern computer techniques should be used for data storage, retrieval, and manipulation.
4. The data collection and entering process should include an appropriate quality control mechanism.

Based on the four criteria the evaluation guidelines of K_{223} can be defined in Table B.3.

Table B.3 scoring guidelines for data reliability

Score / Attribute	1	3	5
Validity K_{223}	$K_{223} = 1$ if none of the validity criteria (illustrated above) is fulfilled	$K_{223} = 3$ if the validity criteria (illustrated above) are partially fulfilled	$K_{223} = 5$ if all of the validity criteria (illustrated above) are fulfilled

Accuracy measures how close the estimated or measured value is compared to the true value. Accuracy is determined by random and systematic errors in the measurements (Popek 2017). Since the data involved in nuclear PRA are mostly related to the number of failures or degradations and are usually collected digitally from different sources, systematic errors in the data are very small. This means that the accuracy of data is primarily determined by the random errors. Since the error margin of the confidence interval is widely accepted as a good indicator of the random errors, it can be used as a measure of the data accuracy. Error factor may be defined based on the upper and lower bounds of confidence interval:

$$error\ factor = \sqrt{\frac{U_l}{L_l}}$$

where U_l and L_l are the upper and the lower bounds of confidence intervals. The accuracy of data is, then, scored based on the value of error factors, following the guidelines in Table B.4.Table B.4 scoring guidelines for data reliability

Table B.4 scoring guidelines for data accuracy

Score / Attribute	1	3	5
Accuracy K_{224}	$K_{224} = 1$ if the error factor is greater than 10	$K_{224} = 3$ if the error factor is between 2-10	$K_{224} = 5$ if the error factor is less or equal to 2

The rest of the "leaf" attributes of the reliability of data are evaluated following the guidelines in Table B.5.

Table B.5 scoring guidelines for data reliability

Score / Attribute	1	3	5
Consistency and relevance K_{222}	$K_{222} = 1$ if the data are not from the same type of power plant, or have different characteristics compared to the system under investigation, e.g., different component or model	$K_{222} = 3$ if the data are from the same power plant with the same type of component and the same characteristics of the system under investigation but from different manufacturers	$K_{221} = 5$ if the data are from the same power plant with the same type of components and the components have the same characteristics and the same manufacturer
Timeliness K_{225}	$K_{225} = 1$ if the data has never been updated	$K_{225} = 3$ if the data has been updated a few years ago (10 years and more)	$K_{225} = 5$ if the data are up-to-date and are updated routinely

Appendix C: Evaluation guidelines for leaf attributes under Understanding of Phenomena ($\mathbf{K_3}$)

Table C.1 Scoring guidelines for Phenomenological understanding's leaf attributes

Score / Attribute	1	3	5
Years of experience (human experience on the phenomenon) K_{31}	$K_{31} = 1$ if the phenomenon is new to human being, and no theories about the phenomenon have been developed yet or the theories are incapable to explain well the phenomenon (e.g. black holes)	$K_{31} = 3$ if the phenomenon has been investigated for moderate years of experience with few theories that are consistent with preexisting ones but still, do not explain holistically the phenomena (e.g. nuclear physics)	$K_{31} = 5$ if the phenomenon has been investigated for a long time and well-established theories have been developed to explain the phenomenon, which have been proved by many evidences (e.g. classical physics)
Number of experts involved in the analysis K_{32}	$K_{32} = 1$ if there is no experts related to this domain (the assessors involved are not expert in this domain) or the experts are unreliable	$K_{32} = 3$ if there is a moderate number of experts of acceptable reliability (two experts) or a low number of experts of high reliability	$K_{32} = 5$ if there is a sufficient number of highly reliable experts (more than two experts)
Academic studies on the phenomena (measured by the number of articles and books published on the subject) K_{33}	$K_{33} = 1$ if no or limited published articles supports the understanding of the phenomenon (e.g. Einstein electromagnetic waves)	$K_{33} = 3$ if a moderate amount of the published articles supports the understanding of the phenomenon (e.g. nuclear energy)	$K_{33} = 5$ if a large amount of the published articles supports the understanding of the phenomenon (e.g. kinetic energy)
Industrial pieces of evidence and applications on the phenomena (measured by the number of applications on available on this subject) K_{34}	$K_{34} = 1$ if no or few industrial applications and reports support the understanding of the phenomenon (e.g. autonomous vehicles)	$K_{34} = 3$ moderate amount of industrial applications and reports support the understanding of the phenomenon (e.g. machine learning)	$K_{34} = 5$ if a large amount of industrial applications and reports support the understanding of the phenomenon (e.g. airplanes)

Appendix V **(P5):**

Tasneem Bani-Mustafa, Zhiguo Zeng, Enrico Zio and Dominique Vasseur "A new framework for multi-hazards risk aggregation" Safety Science (Accepted)

A new framework for multi-hazards risk aggregation

Tasneem Bani-Mustafa [(1)], Zhiguo Zeng [(1)], Enrico Zio [(2) (3) (4)], Dominique Vasseur [(5)]

[(1)] *Chair on System Science and the Energetic Challenge, EDF Foundation*
Laboratoire Genie Industriel, CentraleSupélec, Université Paris-Saclay,
3 Rue Joliot Curie, 91190 Gif-sur-Yvette, France

[(2)] *MINES ParisTech, PSL Research University, CRC, Sophia Antipolis, France*

[(3)] *Energy Department, Politecnico di Milano, Via Giuseppe La Masa 34, Milan, 20156, Italy*

[(4)] *Eminent Scholar, Department of Nuclear Engineering, College of Engineering, Kyung Hee University, Republic of Korea*

[(5)] *EDF R&D, PERICLES (Performance et prévention des Risques Industriels du parc par la simulation et Les Etudes) EDF Lab Paris Saclay - 7 Bd Gaspard Monge, 91120 Palaiseau, France*

Abstract

In this paper, we develop a new method for Multi-Hazards Risk Aggregation (MHRA). A hierarchical framework is first developed for evaluating the trustworthiness of the risk assessment. The evaluation is based on two main attributes (criteria), i.e., the strength of knowledge supporting the assessment and the fidelity of the risk assessment model. These two attributes are further broken down into sub-attributes and, finally, leaf attributes. 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). Risk aggregation is, then, performed by a "weighted posterior" method, considering the level of trustworthiness. An application to the risk aggregation of two hazard groups in Nuclear Power Plants (NPP) is illustrated.

Keywords

Quantitative Risk Assessment (QRA), Risk-Informed Decision Making, Trustworthiness in Risk Assessment, Multi-Hazards Risk Aggregation (MHRA), Strength of Knowledge (SoK), Nuclear Power Plants (NPP)

Acronyms

AHP: Analytical Hierarchy Process
DST: Dempster Shafer Theory
DM: Decision Making
DST-AHP: Dempster Shafer Theory-Analytical Hierarchy Process
EDF: Electricité De France
EFWS: Emergency Feedwater System
IAEA: International Atomic Energy Agency
SoK: Strength of Knowledge
MHRA: Multi-Hazards Risk Aggregation
NPP: Nuclear Power Plants
PRA: Probabilistic Risk Assessment
RIDM: Risk-Informed Decision-Making
USNRC: United-States Nuclear Regulatory Commission

1. Introduction

In Risk-Informed Decision-Making (RIDM), risk metrics are first calculated through Multi-Hazards Risk Aggregation (MHRA) by combining all the relevant information on risk from different contributors (hazard groups) and, then, used to support Decision-Making (DM) (EPRI 2015). A fundamental criticism of the current practice is that the aggregation is conducted by a simple arithmetic summation of the risk metrics from different hazard groups, without considering the heterogeneity in the degrees of maturity and realism of the risk analysis for each hazard group (EPRI 2015). For example, in Nuclear Power Plants (NPP), the Probabilistic Risk Assessment (PRA) for internal events has been developed for many years and considered relatively mature compared to external events (EPRI 2015) or to fire (Siu et al. 2015). Simply adding up the risk indexes can be misleading because it does not consider any information on the trust in the risk indexes calculated for each hazard group. This is a real problem as the results of the PRAs to be aggregated often involve different hazard groups with different levels of realism and trustworthiness.

Various factors contributing to the trustworthiness of risk analysis have been discussed in the literature, including the strength of background knowledge, conservatism, plausibility and realism of assumptions, uncertainty, level of sophistication and details in the analysis, value-ladenness of the assessors, experience, number of approximations and assumptions made in the analysis, etc. (EPRI 2012), (EPRI 2015). Communicating these factors to the decision maker can better inform decision making (Flage & Aven 2009), (EPRI 2012), (Aven 2013b), (EPRI 2015), (Veland & Aven 2015). For this, some experts propose a broad representation of risk that highlights uncertainties rather than probability (Flage & Aven 2009), (Aven, 2013b), (Aven and Krohn, 2014). In Aven (2013a), the risk is described in terms of events, consequences, uncertainty (A, C, U) and a structure is presented for linking the elements of a Data-Information-Knowledge-Wisdom hierarchy to this perspective. In (Flage and Aven, 2009), the authors apply the concept of uncertainty as the main component of risk, whereas the probability is regarded as an epistemic-based expression of uncertainty. Their argument is that for decision making purposes, a broad and comprehensive representation of risk is required to cover the events, consequences, predictions, uncertainty, probability, sensitivity, and knowledge. In addition, they propose a simple and practical method to classify uncertainty factors and evaluate the background knowledge given the following criteria: the inter-alia assumptions and presuppositions (solidity of assumptions), historical field data (availability of reliable data), understanding of phenomena, and agreement among experts.

Some attempts are also found in the literature that focus on treating the uncertain assumptions as an

implication of new risk perspectives. Aven (2013b) proposed a method for assessing the assumption deviation risk by three elements: (i) the degree of the expected deviation of the assumption from reality and its consequences (ii) a measure of uncertainty of the deviation and consequences; (iii) the knowledge on which the assumptions are based. Berner and Flage (2016) summarize four approaches for treating uncertain assumptions: (i) law of total expectation; (ii) interval probability; (iii) crude strength of knowledge and sensitivity categorization; (iv) assumption deviation risk. In this work, they extend the method in Berner and Flage (2015) that evaluates the assumption deviation risk based on three criteria: belief in the deviation from the assumption, sensitivity of the risk index and its dependency on the assumption, and SoK on which the assumptions are made. Six settings are identified for the corresponding scenarios resulting from the three criteria. Guidance for treating the uncertainty related to the deviation of assumptions is given for each setting. Finally, an application of Numeral Unit Spread Assessment Pedigree (NUSAP) is proposed for analyzing the strength, importance, and potential value-ladenness of assumptions through a pedigree diagram. The pedigree diagram uses seven criteria for evaluating the quality of assumptions: (i) plausibility; (ii) inter-subjectivity peers; (iii) inter-subjectivity stakeholders; (iv) choice space; (v) influence of situational limitations; (vi) sensitivity to view and interests of the analyst (vii) and influence on results (Van Der Sluijs et al. 2005), (Boone et al. 2010), (Kloprogge *et al.,* 2011), (De Jong *et al.,* 2012).

In addition, some attempts are found in the literature for directly evaluating the trustworthiness and other relevant quantities. In Bani-Mustafa *et al.* (2017), the trustworthiness of risk assessment models is evaluated through a hierarchical tree that covers the different factors including modeling fidelity, SoK, number of approximations, amount and quality of data, quality of assumptions, number of model parameters, etc. Trustworthiness is also measured in the literature in terms of maturity and credibility. For example, in Model and Simulation (M&S) and information system, a capability maturity model is used to assess the maturity of a software development process in the light of its quality, reliability, and trustworthiness (Paulk et al. 1993). A predictive capability maturity model has been developed to assess the maturity of M&S efforts through evaluating representation and geometric fidelity, physics and material model fidelity, code and solution verification, model validation and uncertainty quantification, and sensitivity analysis (Oberkampf *et al.*, 2007). In (Zeng et al. 2016), a hierarchical framework has been developed to assess the maturity and prediction capability of a prognostic method for maintenance decision making purposes. The hierarchical tree covers different attributes that are believed to affect the prediction

capability of prognostic methods and the trustworthiness of the results. In (Nasa 2013), a framework is proposed for assessing the credibility of M&S through eight criteria: (i) verification; (ii) validation; (iii) input pedigree; (iv) results uncertainty (v) results robustness; (vi) use history; (vii) M&S management; (viii) people qualification. In (Bani-Mustafa *et al.*, 2017), the trust of the model is evaluated based on the level of maturity of the risk assessment model through four main criteria: (i) uncertainty; (ii) knowledge; (iii) conservatism; (iv) sensitivity. The quality of M&S is assured by the American Society of Mechanical Engineers (ASME) through verification and validation (Schwer 2009). Verification is concerned with evaluating the accuracy of the computational model in representing the conceptual and mathematical model, and validation is concerned with evaluating the accuracy of the model in representing reality (Schwer 2009).

As seen from the discussions above, there are a number of works concerned with the realism and trustworthiness of risk assessment. These works, however, discuss the contributors to trustworthiness separately: different frameworks cover different aspects of the trustworthiness based on different terminologies. A unified and complete framework that covers all the factors contributing to trustworthiness is lacking. Besides, the current state of the art only focuses on the evaluation of trustworthiness but does not consider how to integrate the trustworthiness into the results of risk assessment, neither does it show how to aggregate the risk of different contributors with different levels of trustworthiness.

In this work, we define the trustworthiness of risk assessment as a metric that reflects the degree of confidence in the background knowledge that supports the PRA, as well as in the suitability, comprehensiveness and completeness of the PRA model formulation and implementation in a way that reflects, to the best possible, reality. With this, the objective is, then, to provide a new approach for MHRA considering trustworthiness. Compared to the existing works, the contributions of the current work include:

(i) a unified framework is developed for the evaluation of trustworthiness in risk assessment;

(ii) a method is developed to integrate the trustworthiness in the result of the risk assessment of a single hazard group;

(iii) an approach is developed for MHRA considering the trustworthiness of risk assessment.

The rest of this paper is organized as follows. In Section 2, we present a hierarchical framework for assessing the trustworthiness of PRA models and in Section 3 we show how to apply it in practice. In Section 4, we show how to aggregate the risks considering trustworthiness. Section 5 applies the developed methods to a case study from the nuclear industry. Finally, in Section 6, we conclude this paper and discuss

the potential future work.

2. A hierarchical framework for assessing the trustworthiness of a risk model

As illustrated previously, various factors have been discussed in the literature in relation to the trustworthiness of risk assessment. In this paper, we only focus on some of the most relevant factors. For example conservatism, uncertainty, level of sophistication and details in the analysis, experience, number of approximations and assumptions made in the analysis are identified in (EPRI 2012) and (EPRI 2015) as fundamental factors that influence the realism and trustworthiness of a risk analysis. Background knowledge that supports the risk assessment is also widely accepted as an essential contributor to the trustworthiness (Flage & Aven 2009), (Aven 2013a), (Aven 2013b), (EPRI 2012), (EPRI 2015), (Bani-Mustafa *et al.,* 2018). The assumptions that are inevitably made because of incomplete knowledge or for simplifying the analysis (Kloprogge *et al.,* 2011) are considered crucial for the suitability of risk representation and hence, the trustworthiness of its analysis (Boone et al. 2010), (Kloprogge *et al.,* 2011), (De Jong *et al.,* 2012), (Berner & Flage 2016). The conservatism is also identified as a pivotal contributor to the realism, maturity, and trustworthiness of risk assessment (Aven 2016), (Bani-Mustafa *et al.*, 2017). Sensitivity analysis is also needed for a comprehensive description of risk (Flage & Aven 2009), (Bani-Mustafa *et al.*, 2017). Other factors for evaluating the credibility of M&S include verification, validation, input pedigree, result uncertainty, result robustness, use history, M&S management and people qualification (Nasa 2013).

The factors mentioned above are included in the trustworthiness assessment framework proposed in this paper. Other relevant factors are also considered, for a complete representation of trustworthiness. The trustworthiness of risk assessment is defined in this paper as the degree of confidence that the background knowledge is strong enough to support the PRA and that the PRA model is suitable, correctly and robustly made to make the best use of the available knowledge in order to reflect to the best, reality. Obviously, the background knowledge that supports a risk assessment affects significantly the trustworthiness of its results (Flage & Aven 2009), (Aven 2013a), (Aven 2013b), (Bani-Mustafa *et al.,* 2018). However, having a strong background knowledge is not sufficient to ensure the trustworthiness in the results: the fidelity of the modeling should be also verified. This gives rise to a technically adequate and mature model that is known for its high quality and representativeness of reality (Oberkampf *et al.*, 2007), (Nasa 2013), (Zeng et al. 2016). In addition, the modeling process should follow a high quality and thorough application procedure, in order to have trustworthy risk analysis results (IAEA 2006), (Oberkampf *et al.*, 2007),

(Schwer 2009), (Nasa 2013), (Zeng et al. 2016). Hence, the suitability of the selected model and the quality of its application are considered as relevant attributes in the proposed framework. In fact, since the risk metrics are calculated as a result of modeling and simulation, it is intuitive to understand that the trustworthiness of the risk assessment results can be affected by: the suitability of the selected model, the comprehensiveness and correctness of the application of the model, as well as the background knowledge that supports the modeling and analysis. Besides, having results that are highly sensitive to changes in the input is an indication that the assessment is less trustworthy, as the results might be dramatically affected by even a small change in the input parameters and assumptions (Flage & Aven 2009), (Bani-Mustafa *et al.*, 2017). Accordingly, the robustness of the results is regarded as another factor that affects the trustworthiness of risk analysis. In this framework, we use the acronym SoK to represent the strength of the background knowledge that supports the risk assessment and the term "modeling fidelity" to represent the suitability of the selected model, the quality of its application and the robustness of the results, as shown in Figure 1. These two top-level attributes are further decomposed into more tangible sub-attributes.

It should be noted that in general, knowledge includes explicit knowledge, which can be documented and transferred directly, and implicit knowledge, which is possessed by individuals and cannot be documented or transferred directly. The SoK defined in Figure 1 only concerns the explicit knowledge, whereas implicit knowledge is mostly related to the construction and application of the model. Hence, implicit knowledge is viewed as related to the modeling fidelity. The background knowledge is evaluated in Flage and Aven (2009) considering: (i) availability of reliable data; (ii) phenomenological understanding; (iii) quality and plausibility of assumptions; (iv) agreement among peers. In Bani-Mustafa *et al.* (2018), the background knowledge is evaluated by (i) the solidity of assumptions; (ii) the availability of reliable data; (iii) the understanding of phenomena. Each attribute is further broken down into more tangible sub-attributes that define it. For example, the reliability of data is evaluated by its completeness, consistency, validity, accuracy, and timeliness (Bani-Mustafa *et al.,* 2018).

The quality of assumption is evaluated in the literature by different factors. For example, in an application of Numeral Unit Spread Assessment Pedigree (NUSAP), the quality of assumptions is evaluated by (i) plausibility; (ii) inter-subjectivity peers; (iii) inter-subjectivity stakeholders; (iv) choice space; (v) influence situational limitations; (vi) sensitivity to view and interests of the analyst (vii) and influence on results (Van Der Sluijs et al. 2005), (Boone et al. 2010), (Kloprogge *et al.,* 2011). In this paper, we group these factors into three main categories (Bani-Mustafa *et al., 2018*): (i) quality of

assumptions; (ii) value-ladenness; (iii) sensitivity. Value ladenness is, in turn, considered as an independent variable that affects the quality of the assumptions and is evaluated using seven main criteria (i) the personal knowledge; (ii) the sources of information; (iii) the non-biasedness; (iv) the relative independence; (v) the past experience; (vi) the performance measure; (vii) the agreement among peers (Zio 1996), (Bani-Mustafa *et al.,* 2018).

Nevertheless, some of the SoK attributes are more related to the implicit knowledge and affect the construction and formulation of the modeling process and, hence, they are considered under modeling fidelity and not under SoK. For example, the quality and solidity of assumptions are more related to modeling fidelity, since they affect the formulation of the model. Also, since assumptions are made by experts and inevitably affected by their subjectivity, agreement among peers is considered as a sub-attribute under solidity of assumptions.

In this paper, only the availability of reliable data and phenomenological understanding from (Flage & Aven 2009) are considered for evaluating the SoK. As said earlier, the quality and solidity of assumptions are treated under modeling fidelity. Finally, we add another attribute to cover the data and information related directly to the known hazards. The known potential hazards attributes are next broken down into three sub-attributes that cover: the number of documented known hazards, the accident analysis report and the expert's knowledge about the hazards. The data and phenomenological understanding attributes are further broken into sub-attributes and leaf attributes (illustrated in Figure 1) according to the framework proposed in (Bani-Mustafa *et al.,* 2018).

Other factors related to the suitability of the model and quality of application are also found in the literature. Examples of these factors are: conservatism, level of sophistication and details in the analysis, experience, number of approximations and assumptions made in the analysis, sensitivity, results robustness, use history, level of details and verification (Paté-Cornell 1996), (Flage & Aven 2009), (EPRI 2012), (Nasa 2013), (EPRI 2015), (Aven 2016), (Bani-Mustafa *et al.*, 2017). These attributes are allocated in the hierarchy according to their relevance to the modeling fidelity and categorized in three groups, i.e., suitability of selected model, quality of the application and robustness of the results, whereas other attributes have been added to complement the overall framework for the trustworthiness of the risk assessment. The overall hierarchical framework is presented in Figure 1, and detailed definitions of the attributes, sub-attributes and “leaf” attributes are given in Table 1-4.

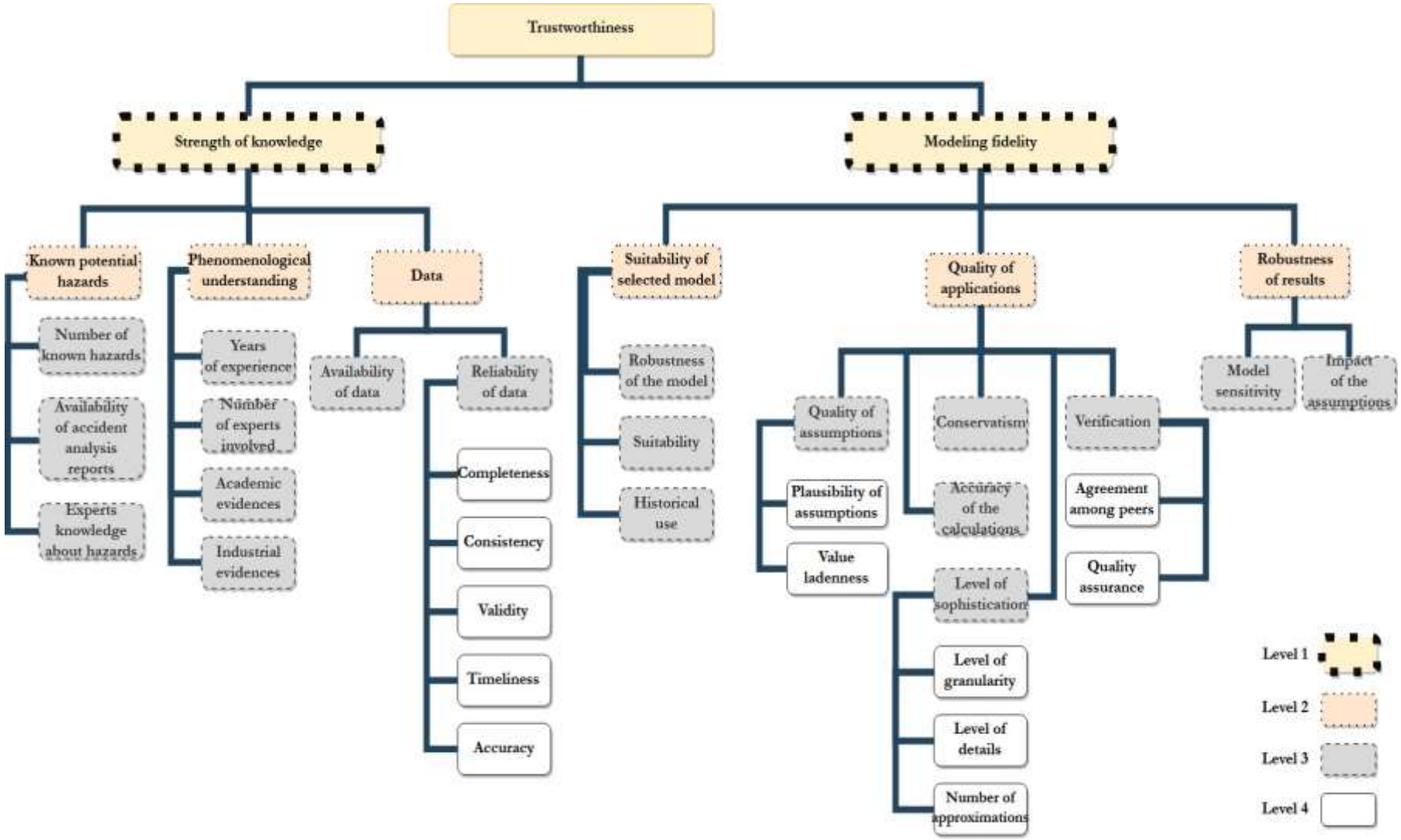


Figure 1 Hierarchical tree for trustworthiness evaluation

Table 1 Definition of trustworthiness attributes (Level 1)

Attribute	**Definition**
Modeling fidelity ($MF = T_1$)	The degree of confidence that the selected PRA model is technically adequate for describing the problem of interest and that the model is implemented in a trustable way so that the results can reasonably represent reality, relative to the decision making involved
The strength of knowledge ($SoK = T_2$)	The amount of high-quality explicit knowledge that is available to support the PRA

Table 2 Definition of trustworthiness attributes (Level 2)

Attribute	**Definition**
Robustness of the results ($RoR = T_{1,1}$)	The capability of the PRA results to remain unaffected by small variations in model parameters or model assumptions
Suitability of the model ($SoM = T_{1,2}$)	The technical adequacy of the tool, maturity and ability to model the problem of interest
Quality of application ($QAp = T_{1,3}$)	The degree to which the analysis is implemented with the minimum required levels of details and modeling adequacy that have the degree of quality, suitable for supporting the application of interest
Knowledge of potential hazards and accident evolution processes ($PoH = T_{2,1}$)	The availability of documentation and knowledge of abnormal events, accidents and their evolutions, from similar systems
Phenomenological understanding ($Ph = T_{2,2}$)	The knowledge that supports the comprehension of the system functionality and the related phenomena
Data ($D = T_{2,3}$)	The amount and quality of data needed for estimating the model parameters

Table 3 Definition of trustworthiness attributes (Level 3)

Attribute	**Definition**
Model sensitivity ($MS = T_{1,1,1}$)	The degree to which the model output varies when one or several parameters change
Impact of assumptions ($IoA = T_{1,1,2}$)	The degree to which the model output varies when one or several assumptions change
Robustness of the model ($RoM = T_{1,2,1}$)	The capability of the model to keep its performance when applied to a different problem settings
Suitability of the model for the problem ($S = T_{1,2,2}$)	The ability to capture all the important details and characterizations of the problem of interest
Historical use ($HU = T_{1,2,3}$)	The degree of confidence gained in this method by the long historical usage
Conservatism ($Cv = T_{1,3,1}$)	The intentional acts for overestimating the risk by making conservative assumptions out of cautiousness
The accuracy of calculations ($AcC = T_{1,3,2}$)	The degree of the voluntarily accepted error in the calculation, e.g., significant figures, simulation errors, and cutoff errors

Quality of assumptions ($QoA = T_{1,3,3}$)	The degree to which the assumption is valid, representing reality and supporting the model
Verification ($Vr = T_{1,3,4}$)	The degree of assurance that the analysis maintains the requirements of quality control standards and obtains the acceptance from different analysts
Level of sophistication ($LoS = T_{1,3,5}$)	The degree of treatment of the problem, and amount of effort and details invested in the problem given its requirement (requirement and complexity)
Number of known hazards ($NH = T_{2,1,1}$)	The documented experience on known hazards that might affect the system of interest
Availability of accident analysis reports ($NH = T_{2,1,2}$)	The availability of technical reports that cover thoroughly the different sequences of any abnormal activity, incident or accident in the time frame and the progressions of each phase
Experts knowledge about the hazard ($NH = T_{2,1,3}$)	The undocumented experience possessed by experts on known hazards
Years of experience ($YE = T_{2,2,1}$)	The amount of experience (measured in years) regarding a specific phenomenon
Number of experts involved ($NE = T_{2,2,2}$)	The number of experts who are explicitly or implicitly involved in understanding the phenomena and the risk analysis
Academic studies on the phenomena ($AE = T_{2,2,3}$)	The number of academic resources, i.e., articles, books, etc., available about the phenomena of interest
Industrial evidence and applications on the phenomena ($IE = T_{2,2,4}$)	The number of industrial applications and reports related to the specific phenomena or events of interest
Amount of available data ($AD = T_{2,3,1}$)	The amount of data that are needed to evaluate the model parameters
Reliability of data ($RD = T_{2,3,2}$)	The degree to which the properties of data satisfy the requirements of risk analysis

Table 4 Definition of trustworthiness attributes (Level 4)

Attribute	**Definition**
The plausibility of assumptions ($Pl = T_{1,3,3,1}$)	The degree of realism of the statements made in the analysis, in cases of lack of knowledge or to facilitate the problem solution
Value ladenness of assessors ($VL = T_{1,3,3,2}$)	The experts' degree of objectivity, professionalism, skills and competencies, past fulfillment of assigned missions and level of achievement
Agreement among peers ($Ag = T_{1,3,4,1}$)	The degree of resemblance between the peers on the analysis and assumptions made, if they were asked to perform the analysis separately
Quality assurance ($QA = T_{1,3,4,2}$)	The degree of following the standards in the process of implementing the analysis
Level of granularity ($LoG = T_{1,3,5,1}$)	The depth of analysis and subdivision of the problem constituting elements
Number of approximations ($NoA = T_{1,3,5,2}$)	The intentional simplifications made to facilitate the modeling
Level of details ($LoD = T_{1,3,5,3}$)	The degree with which the important contributing factors are captured in the modeling compared to the requirement of the analysis (e.g., the dependency among components)
Completeness ($LoD = T_{2,3,2,1}$)	The degree to which the collected data contain the needed information for the risk

	modeling and assessment
Consistency ($LoD = T_{2,3,2,2}$)	The degree of homogeneity of data from different data sources
Validity ($LoD = T_{2,3,2,3}$)	The degree to which the data are collected from a standard collection process and satisfy the syntax of its definition (documentation related)
Timeliness ($LoD = T_{2,3,2,4}$)	The degree to which data correctly reflect the reality of an object or event
Accuracy ($LoD = T_{2,3,2,5}$)	The degree to which data are up-to-date and represent reality for the required point in time

3. Evaluation of the level of trustworthiness

In this section, a bottom-up method for evaluating the level of trustworthiness is developed in Section 3.1. Then, a combination of Dempster Shafer Theory (DST) and Analytical Hierarchy Process (AHP) are used in Section 3.2 to determine the weights of the attributes/sub-attributes in the method proposed in Section 3.1.

3.1. Evaluation of the trustworthiness

In this framework, five levels of trustworthiness are defined with their corresponding settings:

1. 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.

2. 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.

3. 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.

4. 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.

5. 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.

In this paper, the level of trustworthiness of risk assessment is evaluated using a weighted average of the "leaf" attributes in Figure 1.

$$T = \sum_i^n W_i \cdot A_i \tag{1}$$

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; n is the number of the leaf attributes (in Figure 1, we have $n = 27$). The weights W_i are determined based on Dempster Shafer-Analytical Hierarchy Process (DST-AHP) (Dezert *et al.,* 2010), as discussed in Section. 3.2.

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

The weights of the different attributes in Figure1 can be determined using the AHP method to compare their relative importance with respect to the trustworthiness of risk assessment (Saaty 2008). 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, 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, and 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 (Dezert et al. 2010), (Jiao et al. 2016). 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 DST 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 (Dezert et al. 2010), (Dezert & Tacnet 2011), (Jiao et al. 2016). 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. Similar ideas have been used in Tayyebi *et al.* (2010), Ennaceur *et al.* (2011). The procedure for calculating the weights of the leaf attributes based on DST-AHP is presented below.

I. 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 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 is a 3×3 matrix:

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

where the columns correspond to the pairwise comparisons of the daughter attributes: suitability of the selected model, quality of the application, and robustness of the results, 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).

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 (Beynon et al. 2001), (Ennaceur et al. 2011), (Jiao et al. 2016). 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). 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 (Saaty & Vargas 2012).

II. 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 (Saaty 2013). 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.

III. 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 (Jiao et al. 2016). Shafer's reliability discounting is often used to consider the reliability of the source information in DST-AHP (Shafer 1976):

$$m_\delta(A) = \begin{cases} \delta \cdot m(A) & \forall A \subseteq \Theta,\ A \neq \Theta \\ (1-\delta) + (\delta) \cdot m(\Theta), & A = \Theta \end{cases},\ \delta \in [0,1] \tag{2}$$

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.

IV. Combination of experts opinions

Next, Dempster's rule of combination (Shafer 1976) is used to combine two independent pieces of evidence assigned by different experts. The discounted BBAs from different experts are combined by (Jiao et al. 2016):

$$m_{1,2}^{\delta}(C) = \left(m_1^{\delta} \oplus m_2^{\delta}\right)(C) = \begin{cases} 0 & C = \phi, \\ \frac{1}{1-K} \cdot \sum_{A \cap B = C \neq \phi} m_1^{\delta}(A) \cdot m_2^{\delta}(B) & C \neq \phi, \end{cases} \tag{3}$$

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:

$$K = \sum_{A \cap B = \phi} m_1^{\delta}(A) \cdot m_2^{\delta}(B) \tag{4}$$

V. 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 paper, the transferable belief model proposed by (Smets & Kennes 1994) 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 (Smets & Kennes 1994), (Aregui & Denœux 2008):

$$w(x) = \sum_{C \subseteq \Theta, C \neq \phi} \frac{m(C)}{1 - m(\phi)} \frac{1_{C(x)}}{|C|}, \forall x \in \Theta \tag{5}$$

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:

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

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 1, we have $i = 1,2,\cdots,27$.

4. Evaluation of the risk considering trustworthiness levels

In this section, the "weighted posterior" method (Groen & Mosleh 1999) is used for integrating the risk index with the trustworthiness of the PRA for a single hazard group (Section 4.1). In Section 4.2, a structured methodology is developed for determining the weights in the Bayesian "weighted posterior" model. Finally, MHRA considering the level of trustworthiness is discussed in Section 4.3.

4.1. Evaluation of the risk of a single hazard group

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 (Groen & Mosleh 1999). 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:

$$Risk|T = P(E_T) \cdot \text{Risk}|E_T + (1 - P(E_T)) \cdot \text{Risk}|E_{NT} \quad (6)$$

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; $\text{Risk}|E_T$ is the estimated risk from the PRA. Due to the presence of epistemic (parametric) uncertainty in the analysis, $\text{Risk}|E_T$ is often expressed as a subjective probability distribution of the risk index. $\text{Risk}|E_{NT}$ is an alternate distribution of the risk when the decision maker thinks the PRA is not trustable. In this paper, we assume $\text{Risk}|E_{NT}$ 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 (Kaplan & Garrick 1981). For example, Kazemi and Mosleh (2012) developed a similar model to calculate the default risk in similar scenarios considering the unexpected events.

The following steps summarize how to use Eq. (6) to evaluate the risk given the trustworthiness of the risk assessment:

i. The risk distribution $\text{Risk}|E_T$ is evaluated for each hazard group using conventional PRA considering the parametric uncertainty propagation.

ii. The level of trustworthiness of PRA of the corresponding hazard group is assessed, using the procedures in Section 3.

iii. The subjective probability of trusting the PRA is determined by the detailed procedures described

in Section 4.2.

iv. The level of trustworthiness is integrated in the risk using Eq. (6).

4.2. Determining the probability of trusting the PRA

The probability $P(E_T)$ in Eq. (6), 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. (Jenkinson, 2005). 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 (Jenkinson, 2005):

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

ii. 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.

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

iv. 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.

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

vi. 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:

i. The elicitor informs the decision maker about the definition of the different levels of trustworthiness and their physical meaning, based on the definitions in Section 3.1.

ii. 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.

iii. The experts exchange information between them and discuss.

iv. 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.

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

$$p = \frac{x}{1000} \tag{7}$$

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

vi. 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.

The risk assessor can eventually use this function to estimate the probabilities of trust for each hazard group.

4.3. MHRA considering trustworthiness levels

The main steps for MHRA considering trustworthiness are presented in Figure 2. 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), where $Risk_{total}$ is the total risk considering the level of trustworthiness; $(Risk_i|\mathrm{T})$ 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 $Risk_{total}$.

$$Risk_{total} = \sum_{i=1}^{n}(Risk_i|\mathrm{T}) \quad (8)$$

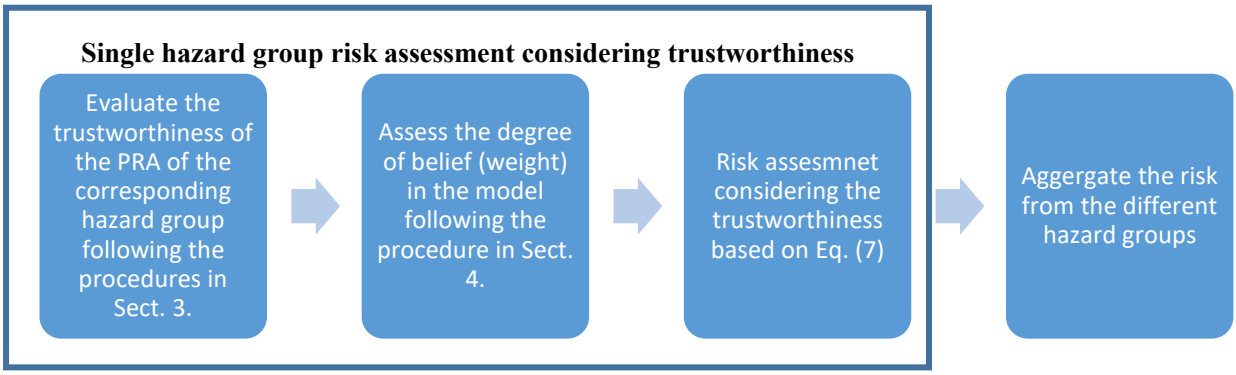


Figure 2 Main steps for MHRA considering the trustworthiness of the PRA

5. Case study

In this section, we apply the developed framework to 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) (Bani-Mustafa *et al.,* 2018). The level of trustworthiness is, then, assessed for each hazard group (Section 5.2). 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 (Section 5.3).

5.1. Description of the PRA model

The two hazard groups considered in this framework are external flooding and internal events. The external flooding refers to the overflow of water that is caused by naturally induced hazards such as river overflows, tsunamis, dam failures and snow melts (IAEA, 2003), (IAEA, 2011). The internal events refer to any undesired event that originates within the NPP and can cause initiating events that might lead to abnormal states and eventually, a core meltdown (EPRI, 2015). Examples of internal events include structural failures, safety systems operation and maintenance errors, etc. (IAEA, 2009). In this case study, the risk analysis is provided by EDF (Bani-Mustafa *et al.,* 2018), in which bow-tie models are used to assess the probability of core damage frequency (CDF). In the original work of EDF, the uncertainty propagation was implemented, but only the mean values of the probability distributions of the risk were considered in MHRA and used for comparison to the safety criteria. However, due to confidentiality reasons, real values cannot be presented. Instead, we disguise the risk distribution, considering also the parametric uncertainty for illustration purposes, as shown in Figure 3.

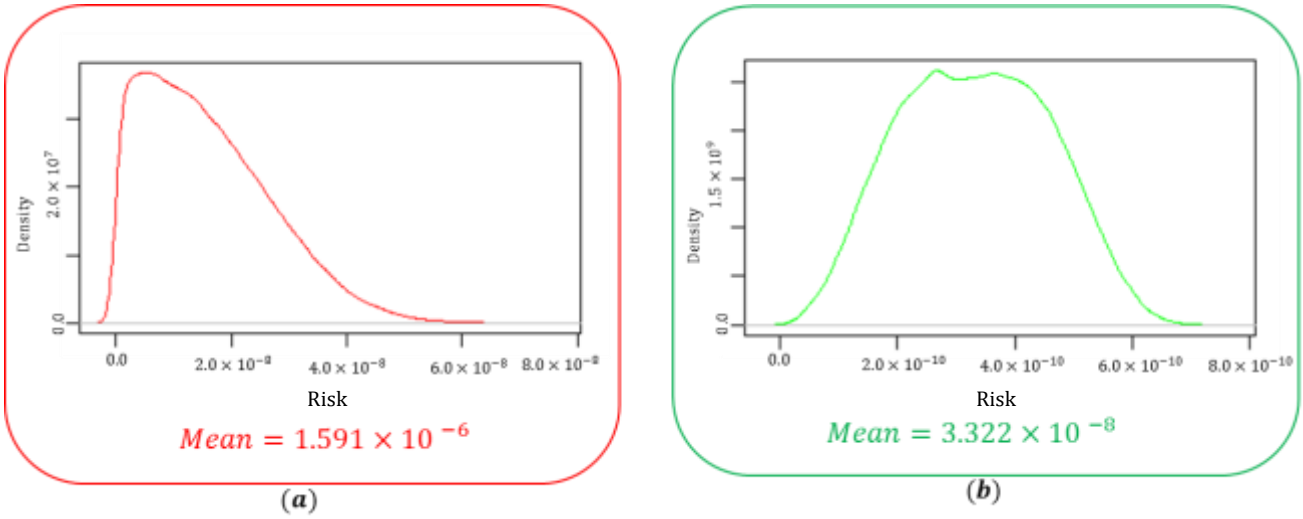


Figure 3 Probability distribution of the risk considering parametric uncertainty: (a) external flooding risk, (b) internal events

5.2. Evaluation of level of trustworthiness

5.2.1. Evaluation of the attributes weights

As illustrated in Section 3, the first step for evaluating the level of trustworthiness is to determine the relative importances (weights) of the trustworthiness attributes. The weights of the attributes are evaluated using the DST-AHP technique. Here, for explanation purposes, the sub-attribute "modeling fidelity" (T_1) is taken as an example to illustrate how to obtain local weights through pairwise comparisons and DTS-AHP.

I. Constructing pairwise comparison matrices

As shown in Section 3, the first step in the DST-AHP technique is to construct the pairwise comparison matrix. Take the daughter attributes of modeling fidelity as an example. In this example, a 4×4 pairwise comparison matrix is constructed in Table 5.

Table 5 Pairwise comparison matrix (knowledge matrix) for comparing modeling fidelity "daughter" attributes

Modeling fidelity	$\{T_{1,1}\}$	$\{T_{1,2}\}$	$\{T_{1,3}\}$	$\Theta = \{T_{1,1}, T_{1,2}, T_{1,3}\}$
$\{T_{1,1}\}$	1	0	0	1/2
$\{T_{1,2}\}$	0	1	0	5/2
$\{T_{1,3}\}$	0	0	1	4
$\{T_{1,1}, T_{1,2}, T_{1,3}\}$	2	2/5	1/4	1

Please note that the zeros that appear in the matrix indicate that there is no need to compare the individual criteria directly: they are compared indirectly through comparing the individual criteria to the universal set Θ (Dezert et al. 2010). $T_{1,1}$ represents the Quality of application, $T_{1,2}$ represents the Suitability of the model, $T_{1,3}$ represents the robustness of the results

In this matrix, the expert has considered four groups of focal sets: three for individual criteria and one

containing all the criteria in order to consider the uncertainty in the evaluation. Choosing focal sets like this means that to the best of their knowledge, the experts believe that the aforementioned focal sets can be favorably compared to the universal set Θ.

II. Computing the weights

In the previous example, the expert was asked to fill the pairwise comparison matrix to express his/her preference of a criterion over another. In this step, the weights of the focal sets are derived using the conventional AHP technique, where the normalized principal eigenvector of the matrix represents the weights. This can be directly done by normalizing each column in the matrix individually and, then, averaging the elements in each row to obtain that weight.

Table 6 Normalized pairwise comparison matrix (knowledge matrix) of modeling fidelity "daughter" attributes

Modeling fidelity	$\{T_{1,1}\}$	$\{T_{1,2}\}$	$\{T_{1,3}\}$	$\Theta = \{T_{1,1}, T_{1,2}, T_{1,3}\}$	Weight (BBA)
$\{T_{1,1}\}$	0.33	0	0	0.06	0.10
$\{T_{1,2}\}$	0	0.71	0	0.31	0.26
$\{T_{1.3}\}$	0	0	0.8	0.5	0.32
$\{T_{1,1}, T_{1,2}, T_{1,3}\}$	0.67	0.29	0.2	0.13	0.32

III. Reliability discounting

After computing the BBA for each expert matrix, the weights need to be discounted based on the reliability of each expert. For illustration purposes, the reliability δ of the expert who made the assessment is assumed to be 0.60. From Eq. (2), the discounted weights are found as the following:

$$m_{0.60}(T_{1,1}) = 0.6 \times 0.10 = 0.06$$

Similarly, for $m_{0.60}(T_{1,2}) = 0.16$, & $m_{0.60}(T_{1,3}) = 0.19$.

Finally, $m_{0.60}(\Theta)$ is found as the following:

$$m_{0.60}(\Theta) = (1 - 0.60) + 0.6 \times 0.32 = 0.59$$

Please note that the BBAs (weights) sum to one before and after the discounting.

IV. Combination of experts opinions

In this case study, three experts have been invited to evaluate the weights; their assigned BBAs are summarized in Table 7 (the BBAs are calculated following the steps in Section 3.2).

Table 7 Discounted basic belief assignment from the three experts

Focal sets of the	Expert 1	Expert 2	Expert 3

criteria	$m_\delta(A)$	$m_\delta(A)$	$m_\delta(A)$
$\{T_{1,1}\}$	0.06	0.16	0.02
$\{T_{1,2}\}$	0.16	0.24	0.38
$\{T_{1,3}\}$	0.19	0.24	0.46
$\{T_{1,1}, T_{1,2}, T_{1,3}\}$	0.59	0.36	0.14

The combination of the experts judgments is conducted sequentially. Table 8 shows the procedures for combining the judgments of the first two experts.

Table 8 Dempster's rule of combination matrix

Expert 2 / Expert 1	$m_\delta(T_{1,1})$	$m_\delta(T_{1,2})$	$m_\delta(T_{1,3})$	$m_\delta(T_{1,1}, T_{1,2}, T_{1,3})$
$m_\delta(T_{1,1})$	$m_\delta(T_{1,1})_1$	ϕ_1	ϕ_2	$m_\delta(T_{1,1})_2$
$m_\delta(T_{1,2})$	ϕ_3	$m_\delta(T_{1,2})_1$	ϕ_4	$m_\delta(T_{1,2})_2$
$m_\delta(T_{1,3})$	ϕ_5	ϕ_6	$m_\delta(T_{1,3})_1$	$m_\delta(T_{1,3})_2$
$m_\delta(T_{1,1}, T_{1,2}, T_{1,3})$	$m_\delta(T_{1,1})_2$	$m_\delta(T_{1,3})_2$	$m_\delta(T_{1,3})_2$	$m_\delta(T_{1,1}, T_{1,2}, T_{1,3})_1$
*Please note that the element ij in the Table represent the multiplication of the elements $1j \times i1$, e.g., $m_\delta(T_{1,1}) \times m_\delta(T_{1,1}) = m_\delta(T_{1.1})_1$; $m_\delta(T_{1,1}) \times m_\delta(T_{1,1}, T_{1,2}, T_{1,3}) = m_\delta(T_{1.1})_2$				

From Eq. (4), $K = 0{,}17$.

From Eq. (3):

$$m_{1,2}^{\delta}(T_{1,3}) = \frac{0{,}26}{1 - 0.17} = 0.31$$

The same steps are repeated for the other mass functions and presented in Table 9. Finally, the new results obtained from the combination of the two experts are further recombined with the BBAs from the third matrix. The results are presented in Table 9.

Table 9 Mass function combinations from the experts

Focal sets of the criteria	Combined mass from experts 1 and 2	Combined mass from experts 1, 2 and 3
	$m_\delta(A)$	
$m_{1,2}^{\delta}(T_{1,1})$	0.15	0.05
$m_{1,2}^{\delta}(T_{1,2})$	0.29	0.40
$m_{1,2}^{\delta}(T_{1,3})$	0.31	0.49
$m_{1,2}^{\delta}(T_{1,1}, T_{1,2}, T_{1,3})$	0.25	0.06

V. Pignistic probability transformation

Then, the pignistic mass function is found by Eq. (5):

$$w_{1,2,3}^{\delta}(T_{1,1}) = m_{1,2,3}^{\delta}(T_{1,1}) + \frac{m_{1,2,3}^{\delta}(T_{1,1}, T_{1,2}, T_{1,3})}{3} = 0.05 + \frac{0.06}{3} = 0.07$$

The steps are repeated for the other mass functions and found to be:

$$w_{1,2,3}^{\delta}(T_{1,2}) = 0.42$$

$$w_{1,2,3}^{\delta}(T_{1,3}) = 0.51$$

Note that the three mass functions on the pignistic level sum to one. These pignistic mass functions represent the relative "believed weights" of the three criteria under modeling fidelity after the reliability discounting and transformation. The same steps are repeated for all the criteria. Then, the weights need to be evaluated with respect to the top-level goal: the trustworthiness. As illustrated previously, this can be done easily by multiplying the weight of the daughter attribute by the weight of the upper parent attributes in each level. For simplicity reasons, only the weights of the "leaf" attribute with respect to the top level attribute i.e., trustworthiness, are presented in Tables 10 and 11 (see Section 5.2.2). Note that the weights of the 27 leaf-attributes with respect to the top goal sum to one $\sum_{i=1}^{27} W_i = 1$.

5.2.2. Evaluation of the attributes scores

The next step is to evaluate the attributes score for the hazard group, given the scoring guidelines in Appendixes A-B. Some information regarding the risk assessment process is extracted from the PRA report to support the trustworthiness assessment:

- The heights (water levels) at the plant's platform at which the water can lead to a failure of a specific element were defined.
- The water flowrate that would result in a given water height at the NPP platform in a defined interval of time was predicted.
- The flow-rate was multiplied by a safety factor of 130%.
- The "return period" for each flowrate was obtained from the data of the millennial flooding flowrate of the river of interest and the data were extrapolated to assess the frequencies of extreme flowrates.
- The river flooding is considered as a predictable phenomenon and the probability of failure of transition into the emergency state (i.e., normal shutdown and cooling with steam generator,

residual heat removal system, etc.) is assumed to be the intrinsic probability of failure.

- It is assumed that river overflow is the only source of external flooding.
- A combined hydraulic/hydrologic method is adopted, given the special hydrological and physical characteristics of the basin.
- It is assumed that once the water reaches the bottom of the equipment, the equipment fails.
- It is assumed that failing to close the valves (ensuring the volumetric protection sealing-water proofing) causes the total loss of Emergency Feedwater System (EFWS).
- It is assumed that clogging inevitably occurs if the flooding occurs.
- The analysis and model calculation for this hazard group is taken with a specific cutoff error of 10^{-14}.

Based on the excerptions from the report, it can be seen that:

- In this example, the risk analysis and assessment steps follow the IAEA recommendations.
- The calculation of flowrates and flow frequencies are calculated using solid deterministic models. However, extrapolation of the data to obtain the frequencies of floods with extreme flowrates is still doubtful.
- The river overflow is a predictable phenomenon and does not happen suddenly. However, the river overflow is not the only source of flooding. For example, a rupture in the river dikes might also lead to sudden, unpredictable flooding.
- The application of a combined hydraulic/hydrologic method on the flooding studies of nuclear sites allows a more realistic evaluation of the flooding level and to estimate more precisely the return periods.
- The assumption that the water will fail the equipment directly if it touches its bottom level is conservative.
- Feedback data show that clogging due to river flooding has occurred before in the nuclear industry (see, for example, USNRC General Electric Advanced Technology Manual for more information (NRC 2011)). However, claiming that each flooding would surely lead to clogging is still questionable and needs to be studied in details, taking into account the different influencing parameters (hydraulic, geometrical and topographical properties) of the area (see (Gschnitz*er et al.*, 2017)).
- In case of failing to close the valves ensuring the volumetric protection, the probability that

water will go back through the drainage system is not identified and assumed to be one ($P = 1$), though there are no relevant calculations. Moreover, once the water enters the physical protection locations, the safety-related equipment is assumed to be lost. Both assumptions are conservative to increase the safety margin.

Based on the above observations, the leaf attributes in Figure 1 can be evaluated. For example, quality assurance attribute is evaluated to be five ($T_{1,3,4,2} = 5$), since the PRA is conducted following the IAEA recommendations. The accuracy of the calculation is evaluated to be five ($T_{1,3,2} = 5$), since the cutoff error is apparently very low. The combined hydraulic/hydrologic models used for the flooding studies are able to capture the special hydrological and physical characteristics of the basin, which makes them suitable for the study. Hence, a score of four ($T_{1,2,2} = 4$) is given for the suitability of the model. The assumptions presented above are mostly conservative and unrealistic. Therefore, a score of one ($T_{1,3,3,1} = 1$) is given for the plausibility of the assumptions. The other attributes are scored in the same way. The results are represented in Tables 10 and 11. The level of trustworthiness for the external flooding is, then, calculated by Eq. (1): $T_{ext} = \sum_{i=1}^{27} W_i \cdot A_i = 3.260$.

Table 10 level-3 leaf attributes weights W and scores S for external flooding hazard group

Att							***Ao C***	***NH***	***AR***	***EK***	***YE***	***NE***	***Ac***	***In***	***AD***
	MS	***IoA***	***RM***	***S***	***HU***	***Cv***									
W	0.012	0.026	0.025	0.158	0.070	0.025	0.012	0.022	0.032	0.054	0.034	0.017	0.105	0.105	0.065
Score	2	2	3	4	3	4	5	2	2	3	3	4	3	3	3

Table 11 level-4 leaf attributes weights W and scores S for external flooding hazard group

Att	***Pl***	***VL***	***Ag***	***QA***	***LoG***	***NoA***	***LoD***	***C***	***Co***	***V***	***T***	***Ac***
W	0.037	0.029	0.025	0.066	0.006	0.005	0.004	0.017	0.011	0.009	0.011	0.017
Score	1	4	4	5	4	4	4	3	3	3	3	3

The trustworthiness for internal events hazard group (T_{int}) was calculated in the same way and, the result is $T_{int} = 4.414$. These results confirm the expectations that the PRA for internal events is considered relatively mature and well established (EPRI 2015) in contrast to the PRA of external hazards, which is considered less mature with several limitations (EPRI 2012).

5.3. Risk assessment considering the level of trustworthiness

5.3.1. Determining the probability of trust in the PRA results

In this step, the decision maker is asked to assign a probability that represents the belief that the risk assessment model output is correct (hereafter called probability of trust), based on the certainty equivalent approach presented in Section 4.2. In this example, we assume that the decision maker exerts a risk-prone behavior and generates the results in Table 12. The data in Table 12 are extrapolated and fitted to a function, as shown in Figure 4.

Table 12 Probability of trust given the level of trustworthiness

Trustworthiness	Probability of trust
1	0.05
2	0.50
3	0.75
4	0.90
5	1.00

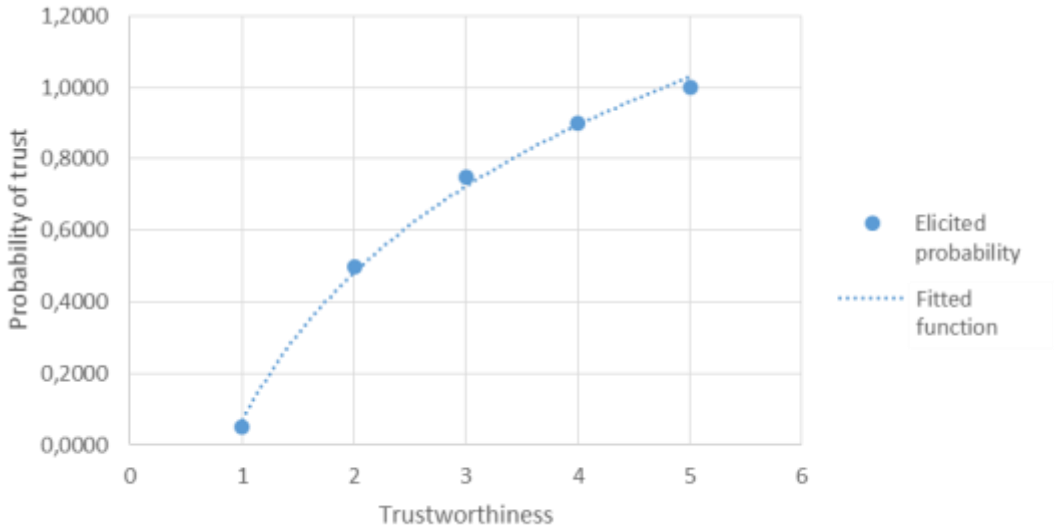


Figure 4 Fitted probability of trust in the PRA given the trustworthiness

Then, the probability that the decision maker trusts each hazard group PRA given their trustworthiness is calculated from the fitted model in Figure 4. The probability of trust for the external flooding p_{ext} is found to be $p_{ext} = 0.783$. The probability of trust for the internal events p_{int} is found to be $p_{int} = 0.957$.

5.3.2. Risk assessment of a single hazard group considering the level of trustworthiness

The level of trustworthiness is integrated with the PRA results for both hazard groups following Eq. (6). The results are presented in Figures 5 and 6, respectively. As illustrated in Figure 5, the mean risk value considering the trustworthiness is 1.088×10^{-1} for external flooding compared to 1.589×10^{-6}

without considering the level of trustworthiness. For internal events, the mean risk value is 2.149×10^{-2} considering the trustworthiness compared to 3.322×10^{-8} without considering it for internal events, as illustrated in Figure 6. It can be seen from the Figures that considering the level of trustworthiness will lead to a larger spread out of the probability distribution of the risk. This comes out as a result of accounting for the disbelief in the risk analysis that reflects the ignorance about the real value of risk. Hence, the spread of the risk distribution becomes wider, leading to a higher mean value of the risk.

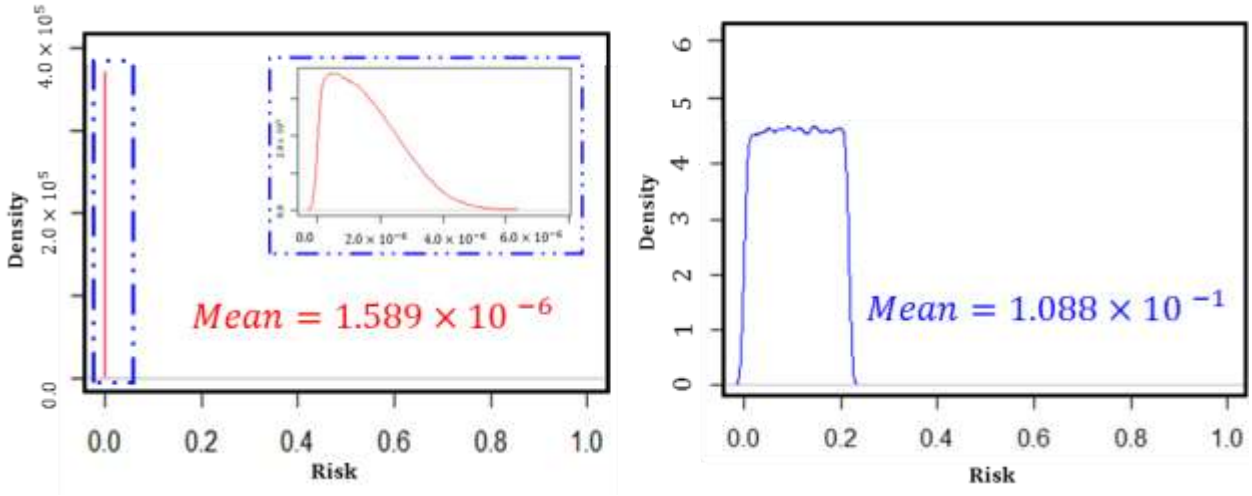


Figure 5 Updated risk estimates after considering the level of trustworthiness for external flooding (a) original risk estimate from the PRA, (b) Risk estimates after integrating the level of trustworthiness

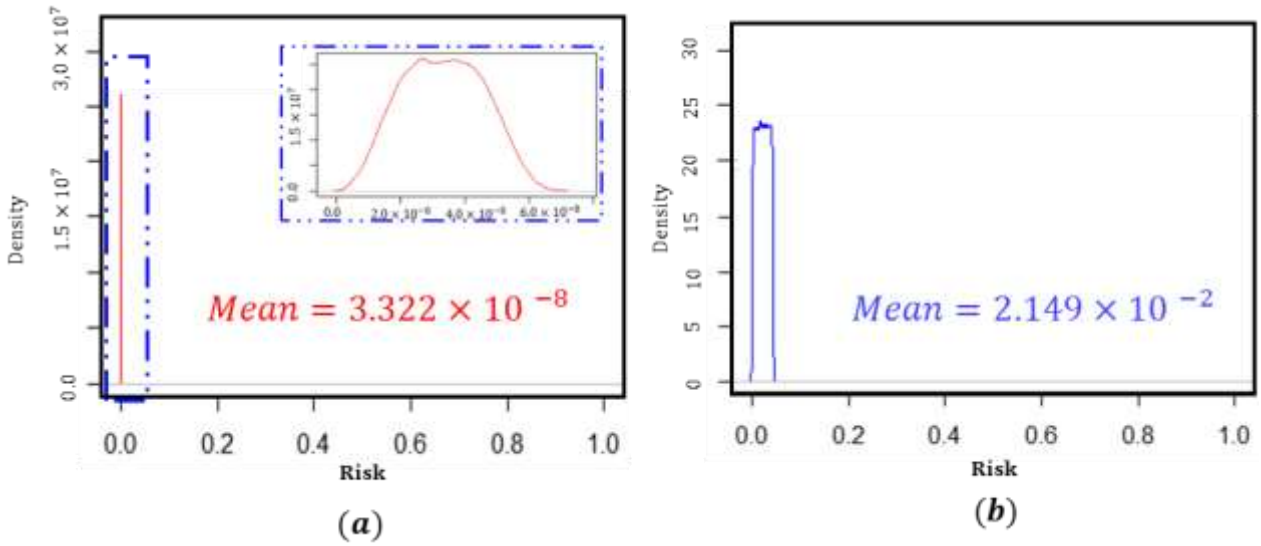


Figure 6 Updated risk estimates after considering the level of trustworthiness for internal events (a) original risk estimate from the PRA, (b) Risk estimates after integrating the level of trustworthiness

5.3.3. Multi-Hazards risk aggregation

Finally, the overall risk given the level of trustworthiness can be calculated using Eq. (8). The results are presented in Figure 7. 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 7 (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} compared to 1.622×10^{-6} without considering the level of trustworthiness. 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. The increase of the spread-out of probability distribution of risk leads to a higher mean value of risk. The aggregation of the risks from the two hazard groups considering the level of trustworthiness results in a more meaningful result, as it takes into account the fact that the PRA model of the two hazard groups is based on different levels of trustworthiness.

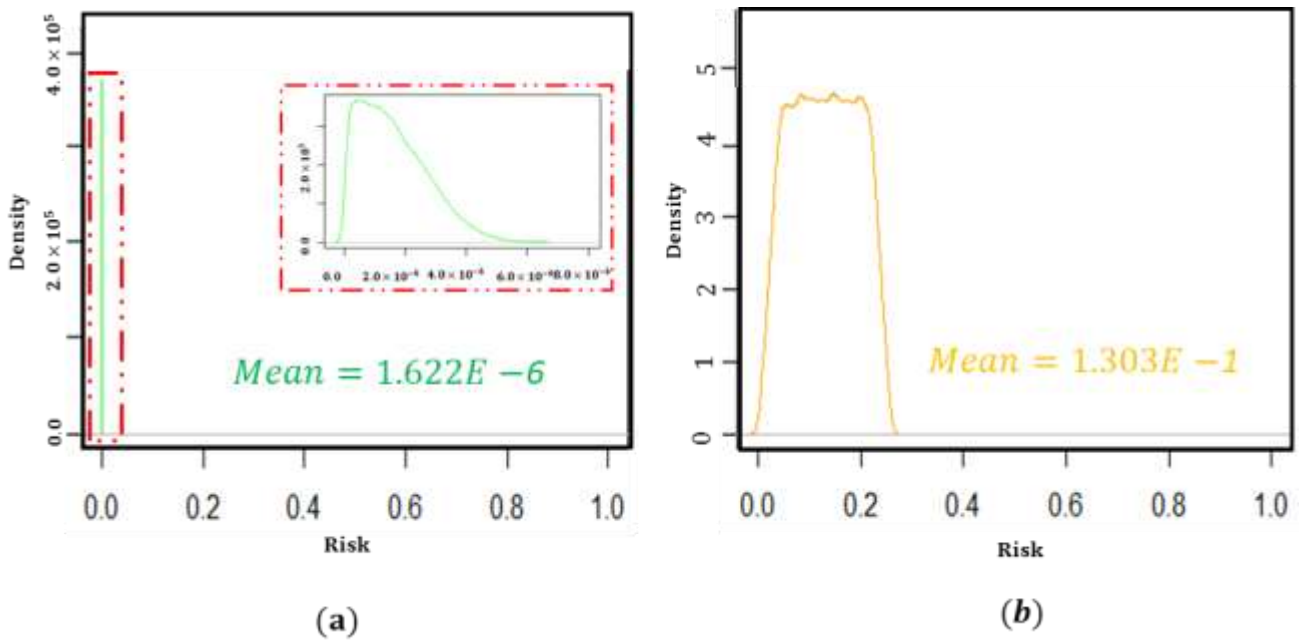


Figure 7 Results of the MHRA, (a) conventional aggregation, (b) considering the level of trustworthiness

6. Discussion and conclusion

In this paper, we have presented a framework for Multi-hazards Risk Aggregation (MHRA) considering trustworthiness. A framework for evaluating the level of trustworthiness is first developed. The

framework consists of two main attributes, i.e., strength of knowledge and modeling fidelity. The strength of knowledge attribute covers the explicit knowledge that can be documented, transferred or explained. The modeling fidelity attribute covers the suitability of the tool and the model construction process. The two attributes are broken down into sub-attributes and, finally, leaf attributes. The total trustworthiness is calculated using a weighted average of the attributes, where the weights are calculated using DST-AHP method.

A MHRA method is, then, developed to aggregate the risk from different hazard groups with different levels of trustworthiness, based on a "weighted posterior" method. An application to a case study of a NPP shows that the developed method allows aggregating risk estimates with different degrees of maturity and realism from different risk contributors.

References

1. Aregui, A. & Denœux, T., 2008. Constructing consonant belief functions from sample data using confidence sets of pignistic probabilities. *International Journal of Approximate Reasoning*, 49(3), pp.575–594.
2. Aven, T., 2013a. A conceptual framework for linking risk and the elements of the data-information-knowledge-wisdom (DIKW) hierarchy. *Reliability Engineering and System Safety*, 111, pp.30–36.
3. Aven, T., 2016. On the use of conservatism in risk assessments. *Reliability Engineering and System Safety*, 146, pp.33–38. Available at: http://dx.doi.org/10.1016/j.ress.2015.10.011.
4. Aven, T., 2013b. Practical implications of the new risk perspectives. *Reliability Engineering and System Safety*, 115, pp.136–145.
5. Bani-mustafa, T. et al., 2018. Strength of Knowledge Assessment for Risk Informed Decision Making. In *Esrel*. Trondheim.
6. Bani-Mustafa, T., Zeng, Z., et al., 2017. A framework for multi-hazards risk aggregation considering risk model maturity levels. In *System Reliability and Safety (ICSRS), 2017 2nd International Conference on*. IEEE, pp. 429–433.
7. Bani-Mustafa, T., Pedroni, N., et al., 2017. A hierarchical tree-based decision making approach for assessing the trustworthiness of risk assessment models. In *PSA (ANS)*. American Nuclear Society (ANS).
8. Berner, C. & Flage, R., 2016. Strengthening quantitative risk assessments by systematic treatment of uncertain assumptions. *Reliability Engineering and System Safety*, 151, pp.46–59.

9. Beynon, M., Cosker, D. & Marshall, D., 2001. An expert system for multi-criteria decision making using Dempster Shafer theory. *Expert Systems with Applications*, 20(4), pp.357–367. Available at: http://www.sciencedirect.com/science/article/pii/S0957417401000203.
10. Boone, I. et al., 2010. NUSAP: a method to evaluate the quality of assumptions in quantitative microbial risk assessment. *Journal of Risk Research*, 13(3), pp.337–352. Available at: http://www.scopus.com/inward/record.url?eid=2-s2.0-77951165131&partnerID=40&md5=12a3caae6ff5f3fae9967becb6b35f17.
11. Dezert, J. et al., 2010. Multi-criteria decision making based on DSmT-AHP. In *BELIEF 2010: Workshop on the Theory of Belief Functions*. Belief Functions and Applications Society (BFAS), p. 8–p.
12. Dezert, J. & Tacnet, J.-M., 2011. Evidential reasoning for multi-criteria analysis based on DSmT-AHP. *Advances and Applications of DSmT for Information Fusion*, p.95.
13. Ennaceur, A., Elouedi, Z. & Lefevre, E., 2011. Handling partial preferences in the belief AHP method: Application to life cycle assessment. In *Congress of the Italian Association for Artificial Intelligence*. Springer, pp. 395–400.
14. EPRI, 2015. *An Approach to Risk Aggregation for RisK-Informed Decision-Making*, Palo Alto, California.
15. EPRI, 2012. *Practical Guidance on the Use of Probabilistic Risk Assessment in Risk-Informed Applications with a Focus on the treatment of Uncertanity*, Palo Alto, California.
16. Flage, R. & Aven, T., 2009. Expressing and communicating uncertainty in relation to quantitative risk analysis. *Reliability: Theory & Applications*, 4(2–1 (13)).
17. Groen, F. & Mosleh, A., 1999. Behavior of weighted likelihood and weighted posterior methods for treatment of uncertain data. In *Proc. ESREL*.
18. Gschnitzer, T. et al., 2017. Towards a robust assessment of bridge clogging processes in flood risk management. *Geomorphology*, 279, pp.128–140. Available at: http://www.sciencedirect.com/science/article/pii/S0169555X1631042X.
19. IAEA, 1991. *Data Collection and Record Keeping for the Management of Nuclear Power Plant Ageing* IAEA, ed.,
20. IAEA, 2006. *Determining the Quality of Probabilistic Safety Assessment (PSA) for Applications in Nuclear Power Plants*, Vienna: INTERNATIONAL ATOMIC ENERGY AGENCY. Available at:

http://www-pub.iaea.org/books/IAEABooks/7546/Determining-the-Quality-of-Probabilistic-Safety-Assessment-PSA-for-Applications-in-Nuclear-Power-Plants.

21. IAEA, 2003. *External Events Excluding Earthquakes in the Design of Nuclear Power Plants*,
22. IAEA, 2011. IAEA-Publication8635.
23. IAEA Safety Standards Series, 2009. *Deterministic Safety Analysis for Nuclear Power Plants*,
24. Jenkinson, D., 2005. *The elicitation of probabilities: A review of the statistical literature*, Citeseer.
25. Jiao, L. et al., 2016. Combining sources of evidence with reliability and importance for decision making. *Central European Journal of Operations Research*, 24(1), pp.87–106.
26. De Jong, A., Wardekker, J.A. & Van der Sluijs, J.P., 2012. Assumptions in quantitative analyses of health risks of overhead power lines. *Environmental science & policy*, 16, pp.114–121.
27. Kaplan, S. & Garrick, B.J., 1981. On the quantitative definition of risk. *Risk analysis*, 1(1), pp.11–27.
28. Kazemi, R. & Mosleh, A., 2012. Improving default risk prediction using Bayesian model uncertainty techniques. *Risk Analysis: An International Journal*, 32(11), pp.1888–1900.
29. Kloprogge, P., Van der Sluijs, J.P. & Petersen, A.C., 2011. A method for the analysis of assumptions in model-based environmental assessments. *Environmental Modelling and Software*, 26(3), pp.289–301. Available at: http://dx.doi.org/10.1016/j.envsoft.2009.06.009.
30. Nasa, 2013. STANDARD FOR MODELS AND SIMULATIONS-NASA-STD-7009. , (I), pp.7–11.
31. NRC, U.S., 2011. *General Electric Advanced Technology Manual Chapter 4.8 Service Water System Problems*,
32. Oberkampf, W.L., Pilch, M. & Trucano, T.G., 2007. Predictive capability maturity model for computational modeling and simulation. *cfwebprod.sandia.gov*. Available at: https://cfwebprod.sandia.gov/cfdocs/CCIM/docs/Oberkampf-Pilch-Trucano-SAND2007-5948.pdf%5Cnfile:///Users/markchilenski/Documents/Papers/2007/cfwebprod.sandia.gov%0A/Oberkampf/cfwebprod.sandia.gov%0A 2007 Oberkampf.pdf%5Cnpapers://31a1b09a-25a9-4e20-879d-4.
33. Paté-Cornell, M.E., 1996. Uncertainties in risk analysis: Six levels of treatment. *Reliability Engineering & System Safety*, 54(2), pp.95–111.
34. Paulk, M.C. et al., 1993. Capability Maturity Model for Software, Version 1.1. *Software, IEEE*, 98(February), pp.1–26. Available at: http://www.sei.cmu.edu/library/abstracts/reports/93tr024.cfm.
35. Popek, E.P., 2017. *Sampling and analysis of environmental chemical pollutants: a complete guide*, Elsevier.

36. Saaty, T.L., 2013. Analytic hierarchy process. In *Encyclopedia of operations research and management science*. Springer, pp. 52–64.
37. Saaty, T.L., 2008. Decision making with the analytic hierarchy process. *International Journal of Services Sciences*, 1(1), p.83.
38. Saaty, T.L. & Vargas, L.G., 2012. *Models, methods, concepts & applications of the analytic hierarchy process*, Springer Science & Business Media.
39. Schwer, L.E., 2009. Guide for Verification and Validation in Computational Solid Mechanics.
40. Shafer, G., 1976. *A mathematical theory of evidence*, Princeton university press.
41. Siu, N. et al., 2015. FIRE PRA MATURITY AND REALISM: A DISCUSSION AND SUGGESTIONS FOR IMPROVEMENT.
42. Van Der Sluijs, J.P. et al., 2005. Combining Quantitative and Qualitative Measures of Uncertainty in Model-Based Environmental Assessment: The NUSAP System. *Risk Analysis*, 25(2), pp.481–492. Available at: http://doi.wiley.com/10.1111/j.1539-6924.2005.00604.x.
43. Smets, P. & Kennes, R., 1994. The transferable belief model. *Artificial intelligence*, 66(2), pp.191–234.
44. Tayyebi, A.H. et al., 2010. Combining multi criteria decision making and Dempster Shafer theory for landfill site selection. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci*, 38(8), pp.1073–1078.
45. Veland, H. & Aven, T., 2015. Improving the risk assessments of critical operations to better reflect uncertainties and the unforeseen. *Safety Science*, 79, pp.206–212. Available at: http://www.sciencedirect.com/science/article/pii/S092575351500154X.
46. Zeng, Z. et al., 2016. A hierarchical decision-making framework for the assessment of the prediction capability of prognostic methods. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 231(1), pp.36–52. Available at: http://dx.doi.org/10.1177/1748006X16683321.
47. Zio, E., 1996. On the use of the analytic hierarchy process in the aggregation of expert judgments. *Reliability Engineering and System Safety*, 53(2), pp.127–138.

Appendix A: Evaluation guidelines for leaf attributes under modeling fidelity (T_1)

Appendix A.1: Attributes under "robustness of the results attributes"

Table A.1.1 Scoring guidelines for robustness of the results

Attribute \ Score	1	3	5
Model sensitivity T_{111}	$T_{111} = 1$ if the ensemble of model parameters greatly influence the final result	$T_{111} = 3$ if the ensemble of model parameters moderately influence the results	$T_{111} = 5$ if the ensemble of model parameters have little or no impact on the results of risk analysis
Impact of the assumptions T_{112}	$T_{112} = 1$ if the assumption greatly influences the results of risk analysis	$T_{112} = 3$ if the assumption moderately influences the results of risk analysis	$T_{112} = 5$ if the assumption has little or no impact on the results of risk analysis

Appendix A.2: Attributes under "suitability of the selected model"

Table A.2.1 Scoring guidelines for suitability of the selected model

Attribute \ Score	1	3	5
Robustness of the model T_{121}	$T_{111} = 1$ if the model doesn't show the capability of performing under different settings or when exerting, deliberately, some variations in the assumptions and parameters	$T_{111} = 3$ if the model show the capability of performing moderately under different settings or small deliberate variations in the assumptions and parameters	$T_{111} = 5$ if the model show the capability of performing under different settings or when exerting, deliberately, large variations in the assumptions and parameters
Suitability of the tool T_{122}	$T_{122} = 1$ if the selected model is not usually used for achieving objectives similar to the required ones or it is not suitable for the problem settings and cannot capture all the important	$T_{122} = 3$ if the selected model is usually used for achieving objectives similar to the required ones or it is suitable for the problem settings but doesn't capture entirely the important aspects of	$T_{122} = 5$ if the selected model is usually used for achieving objectives similar to the required ones and it is suitable for the problem settings in a way that captures entirely the important

	aspects of the problem	the problem	aspects of the problem in a way that makes it suitable to represent reality
Historical use T_{123}	$T_{123} = 1$ if the selected tool is new or has never proved its successful use before, or if it is a new version of the tool that is quite different from the old one	$T_{123} = 3$ if the selected tool is a new updated version of a tool that has proved its successful use before	$T_{123} = 5$ if the selected tool is quite common tool that has proved its successful use in different problem settings, or if it is a slightly updated version of an old common one that proved it successful use

Appendix A.3: Attributes under "quality of application"

Conservatism:

In this setting, the conservatism is evaluated in the light of three criteria: (i) types of risk index estimates (best judgment, true value with a high confidence and true value with a low confidence); (ii) context of decision making; (iii) the effect of conservatism on the perception of the problem compared to best or true estimates or true and consequently decision making assumptions and parameters. Figure A.1-3 illustrate the different score for each corresponding scenario.

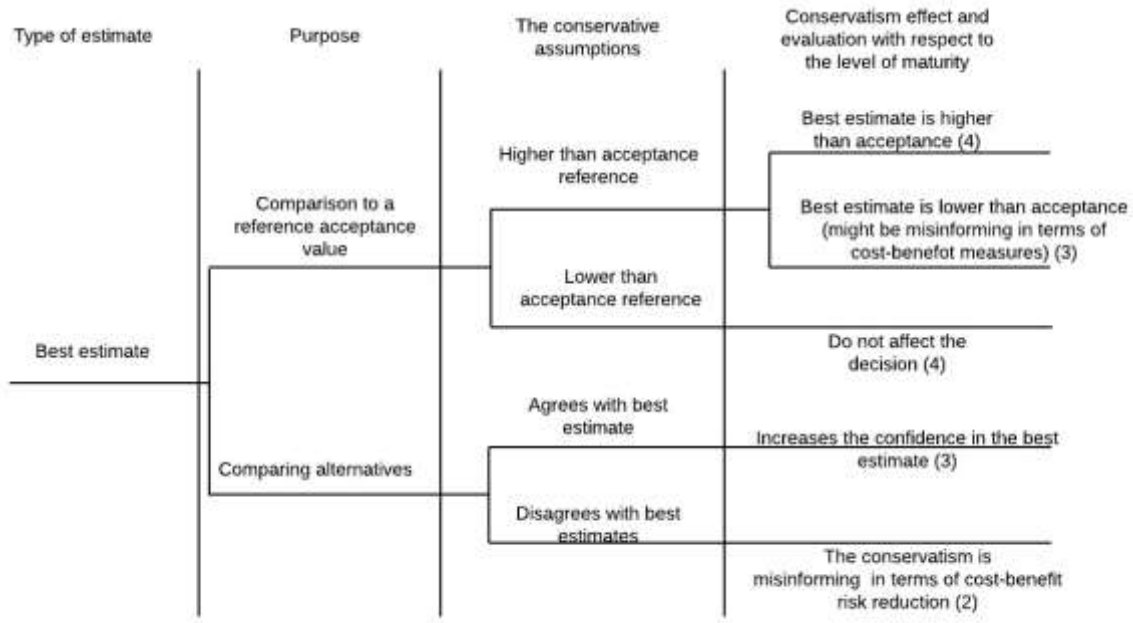


Figure A.3.1 Evaluation of the conservatism in the light of the level of maturity (conservatism VS Best estimate)

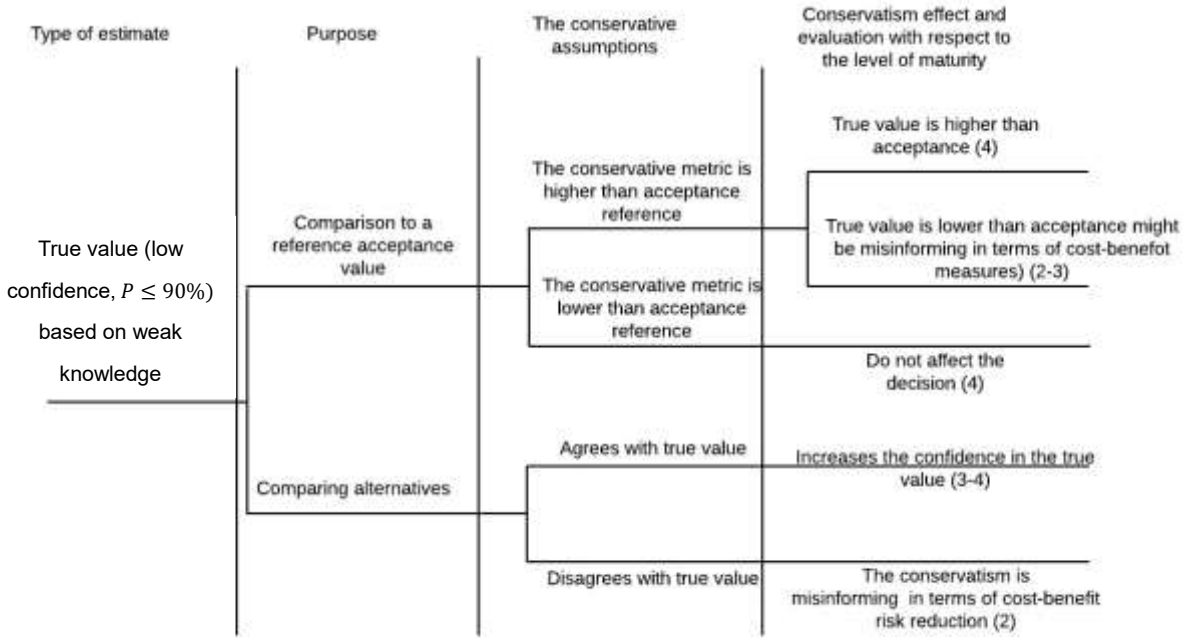


Figure A.3.2 Evaluation of the conservatism in the light of the level of maturity (conservatism VS True value/weak knowledge)

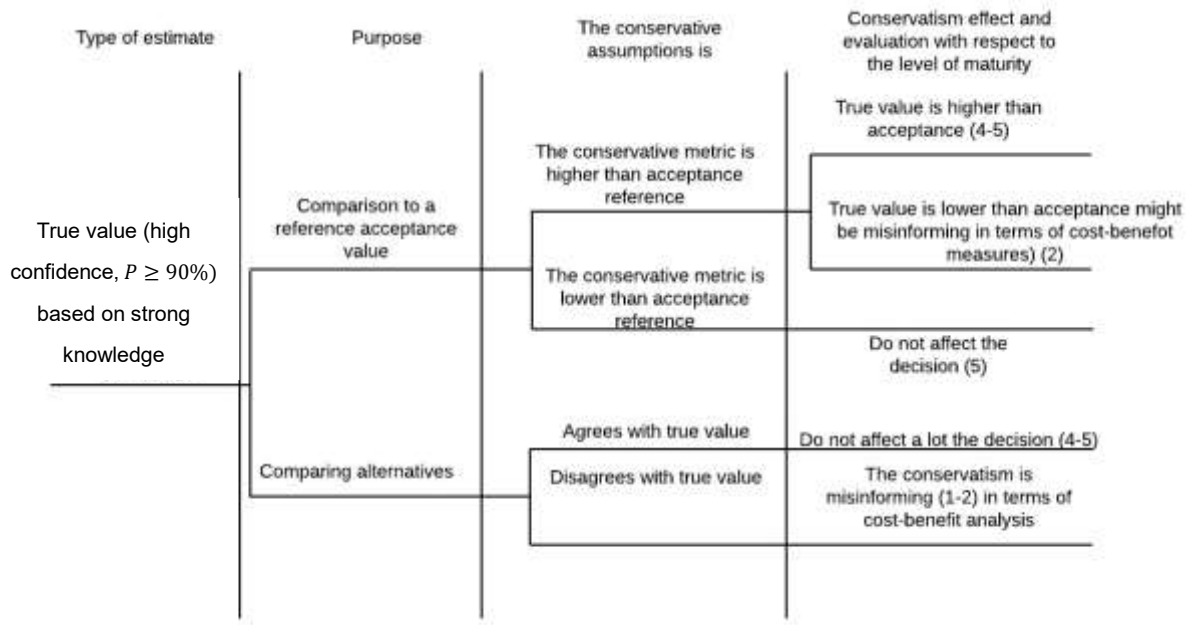


Figure A.3.3 Evaluation of the conservatism in the light of the level of maturity (conservatism VS True value/strong knowledge)

Table A.3.1 Scoring guidelines for the quality of the application

Attribute \ Score	1	3	5
The accuracy of the calculation T_{132}	$K_{131} = 1$ if the setting of accuracy is chosen to be low and high degree of error is accepted in the calculations. For example, the cutoff error (the chosen value of parameters at which lower values are ignored) is set to be large, and a low number of trials are performed	$K_{131} = 3$ if the setting of accuracy is chosen to be acceptable with a tolerable degree of errors. For example, the cutoff error is set to be quite low and a sufficient number of trials are performed	$K_{131} = 5$ if the setting of accuracy is chosen to be high and errors are conservatively accepted in the calculations. For example, the cutoff error is set at to be small, and a high number of trials are performed

Table A.3.2 Scoring guidelines for quality of assumptions *(*Boone *et al.,2010)*

Attribute \ Score	1	3	5

Plausibility of assumptions T_{1331}	$K_{1331} = 1$ if the assumption is not realistic (over conservative or over optimistic), or the available information is not sufficient for assessing the quality of the assumptions	$K_{1331} = 3$ if the assumption is based on existing simple models and extrapolated data	$K_{1331} = 5$ if the assumption is plausible: it is grounded on well-established theory or abundant experience on similar systems, and verified by peer review

Note: If multiple assumptions are involved in the assessment, the final score for T_{1331} is obtained by averaging the scores of all the assumptions.

Table A.3.3 Scoring guidelines for the value-ladenness of the assessors

Score / Attribute	1	3	5
Personal knowledge (educational background) T_{13321}	$T_{13321} = 1$ if all of the experts hold academic degrees from other domains	$T_{13321} = 3$ if less than two thirds of the experts hold academic degrees in the same field	$T_{13321} = 5$ if over two thirds of the experts hold academic degrees in the same field
Sources of information T_{13322}	$T_{13322} = 1$ if experts can only access academic information source or only industrial information source	$T_{13322} = 3$ if experts can access fully industrial information source and partially academic information source	$T_{13322} = 3$ if experts can fully access both academic and industrial information sources
Unbiasedness and plausibility T_{13323}	$T_{13323} = 1$ if the expert team is very conservative or optimistic	$T_{13323} = 3$ if the expert team is slightly conservative/optimistic	$T_{13323} = 5$ if as a team, the experts are unbiased: the biases of the experts can compensate one another
Relative independence T_{13324}	$T_{13324} = 1$ if over three quarters of the experts are highly influenced by mangers and stakeholders	$T_{13324} = 3$ if less than one quarter of experts might be influenced by the mangers and stakeholders	$T_{13324} = 5$ if all experts' decisions are highly independent
Past experience T_{13325}	$T_{13325} = 1$ if the experts' experience is less than 5 years	$T_{13325} = 3$ if the experts' experience is between 10-15 years	$T_{13325} = 5$ if the experts' experience is more than 20 years

Performance measure T_{13326}	$T_{13326} = 1$ if the performance of the experts are not evaluated by external peers	$T_{13326} = 3$ if the external peers generally acknowledge the experts' performance but raise some slight concerns	$T_{13326} = 5$ if the external peers endorse the experts' performance and approve them
*Please note the value-ladenness score is calculated by averaging the scores over all the attributes in this table.			

Table A.3.4 Scoring guidelines for leaf attributes under verification

Score / Attribute	1	3	5
Agreement among peers T_{1341}	$T_{1341} = 1$ if some experts hold strongly conflicting views on the assumptions	$T_{1341} = 3$ if some experts questions on the assumptions, but do not have strongly conflicting views	$T_{1341} = 1$ if most of the experts agree on the assumptions
Quality assurance T_{1342}	$T_{1341} = 1$ if the analysis does not follow the quality standards and recommendations set by the PSA community e.g., ASME standards, NRC regulatory guides, IAEA recommendations	$T_{1341} = 3$ if the analysis follows moderately the quality standards and recommendations set by the PSA community e.g., ASME standards, NRC regulatory guides, IAEA recommendations	$T_{1341} = 5$ if the analysis follows entirely and conservatively the quality standards and recommendations set by the PSA community e.g., ASME standards, NRC regulatory guides, IAEA recommendations

Table A.3.5 Scoring guidelines for leaf attributes under the level of sophistication

Score / Attribute	1	3	5
Level of granularity T_{1351}	$T_{1341} = 1$ if the level of analysis is performed abstractly and coarsely on the level of systems or level the level of large components	$T_{1341} = 3$ if the analysis is performed in to a sufficiently fine level that regards the small components of a system or a small factors of a problem	$T_{1341} = 1$ if the level of analysis is zoomed in to the level of component's small constituting parts e.g., considering the small constituting parts of a

			manual (i.e., valve, the body, bonnet, ports etc.) when building the physical model for calculating the failure rate of a manual valve
Number of approximations T_{1352}	$T_{1342} = 1$ if there is a large number of approximations and the aggregate of the approximations affects significantly the output	$T_{1342} = 3$ if there is a moderate number of approximations or the aggregate of the approximations affects moderately the output	$T_{1342} = 5$ if there is a low number of approximations and the aggregate of the approximations does not affect, or affects insignificantly the output
Level of details T_{1353}	$T_{1353} = 1$ if most of the relevant contributing factors (including those that are not evident in the model construction requirements) that affect the estimates are not captured in modeling process compared to a complete realistic modeling e.g., the dependency among components in calculating the failure of a given component, environmental and thermal effect on components, level of the PH	$T_{1353} = 3$ if most of the relevant contributing factors (including those that are not evident in the model construction requirements) that estimates are captured in the modeling process compared to a complete realistic modeling e.g., considering the dependency among components in calculating the failure of a given component, environmental and thermal effect on components, level of the PH	$T_{1353} = 3$ if all relevant contributing factors (including those that are not evident in the model construction requirements) that affect the estimates are captured in modeling process compared to a complete realistic modeling e.g., considering the dependency among components in calculating the failure of a given component, environmental and thermal effect on components, level of the PH

Appendix B: Evaluation guidelines for the strength of knowledge (T_2) leaf attributes

Appendix B.1: Attributes under "Known potential hazards"

Table B.1.1 Scoring guidelines for leaf attributes under known potential hazards

Score / Attribute	1	3	5
Number of known hazards T_{211}	$T_{211} = 1$ if there is only a few number of known relevant hazards that are considered in the analysis	$T_{211} = 3$ if there is a moderate number of known relevant hazards that are considered in the analysis	$T_{211} = 5$ if there is a high number of known relevant hazards that are considered in the analysis
Availability of accident reports T_{212}	$T_{212} = 1$ if there is no past experience and technical reports that explain and cover in details the timing, causes and different sequences of abnormal activities, incident or accident	$T_{212} = 3$ if there is only a few past experience and technical reports that explain and cover in details the timing, causes and different sequences of abnormal activities, incident or accident, or if there is abundancy of reports that covers accidents without details	$T_{212} = 5$ if there is abundancy of past experience and technical reports that explain and cover in details the timing, causes and different sequences of abnormal activities, incident or accident
Experts knowledge about hazards T_{213}	$T_{213} = 1$ if the expert has a low experience in such a type of analysis and hazards, as well as other types of problem, in a way that prevents him from imagining new unknown types of hazards	$T_{213} = 3$ if the expert has a moderate degree of experience in such a type of analysis and hazards, as well as other types of problem, in a way that allows him to imagine new unknown types of hazards	$T_{213} = 5$ if the expert has a high degree of experience in such a type of analysis and hazards, as well as other types of problem, in a way that allows him to imagine most of the unknown types of hazards

Appendix B.2: Attributes under "phenomenological understanding"

Table B.2.1 Scoring guidelines for phenomenological understandings' leaf attributes

Score / Attribute	1	3	5

Years of experience (human experience on the phenomenon) T_{221}	$T_{221} = 1$ if the phenomenon is new to a human being, and no theories about the phenomenon have been developed yet or the theories are incapable to explain well the phenomenon (e.g., black holes)	$T_{221} = 3$ if the phenomenon has been investigated for moderate years of experience with few theories that are consistent with preexisting ones but still, do not explain holistically the phenomena (e.g., nuclear physics)	$T_{221} = 5$ if the phenomenon has been investigated for a long time and well-established theories have been developed to explain the phenomenon, which have been proved by many evidences (e.g., classical physics)
Number of experts involved in the analysis T_{222}	$T_{222} = 1$ if there is no experts related to this domain (the assessors involved are not expert in this domain) or the experts are unreliable	$T_{222} = 3$ if there is a moderate number of experts of acceptable reliability (two experts) or a low number of experts of high reliability	$T_{222} = 5$ if there is a sufficient number of highly reliable experts (more than two experts)
Academic studies on the phenomena (measured by the number of articles and books published on the subject) T_{223}	$T_{223} = 1$ if no or limited published articles supports the understanding of the phenomenon (e.g., Einstein electromagnetic waves)	$T_{223} = 3$ if a moderate amount of the published articles supports the understanding of the phenomenon (e.g., nuclear energy)	$T_{223} = 5$ if a large amount of the published articles supports the understanding of the phenomenon (e.g., kinetic energy)
Industrial pieces of evidence and applications on the phenomena (measured by the number of applications available on this subject) T_{224}	$T_{224} = 1$ if no or few industrial applications and reports support the understanding of the phenomenon (e.g., autonomous vehicles)	$T_{224} = 3$ moderate amount of industrial applications and reports support the understanding of the phenomenon (e.g., machine learning)	$T_{224} = 5$ if la arge amount of industrial applications and reports support the understanding of the phenomenon (e.g., airplanes)

Appendix B.3: Evaluation guidelines for leaf attributes under "Data"

Amount of data T_{231} is measured by a numerical metric, Years of Experience (YoE), defined by the

number of related events recorded during a specific period.

YoE =length of the data collection period (in years) × sample size of the data

The amount of data is scored based on the criteria in Table B.3.1.

Table B.3.1 Scoring guidelines for Amount of available data

Value of YoE	Score
< 50	1
50-199	2
200-499	3
500-999	4
>1000	5

Completeness of data refers to the degree to which the collected data contains the needed information. For components and systems, data completeness is characterized by the following criteria (IAEA 1991):

1. The data should contain baseline information, which covers the design data and conditions of a component at its initial state.
2. The data should contain the operating history, which covers the service conditions of systems and components including transient and failure data.
3. The data should contain the maintenance history data, which covers the components monitoring and maintenance data.

For more details on how each of the previous attributes is identified, see (IAEA 1991). However, it should be noted that the completeness features are defined differently depending on the problem. For example, data required for quantifying to a component failure frequency is different from that for quantifying a natural event. General scoring guidelines for evaluating T_{2321} are given, based on the degree to which criteria are satisfied, as shown in Table B.3.2.

Table B.3.2 scoring guidelines for data reliability

Score / Attribute	1	3	5
Completeness T_{2321}	$T_{2321} = 1$ if the data fail to contain the necessary information required in developing the risk	$T_{2321} = 3$ if the data contain to an acceptable degree the necessary information required in developing the	$T_{2321} = 5$ if the data contain all the necessary information required in developing the risk assessment model (in the

	assessment model (in the light of the completeness characteristics defined above)	risk assessment model (in the light of the completeness characteristics defined above)	light of the completeness characteristics defined above)

The validity of data is evaluated by the following criteria:

1. The integrity of data is carefully managed.
2. Databases are well organized and formatted in a common way, and easily retrieved and manipulated.
3. Data should be collected and entered in the database by well-trained maintenance personnel, and modern computer techniques should be used for data storage, retrieval, and manipulation.
4. The data collection and entering process should include an appropriate quality control mechanism.

Based on the four criteria the evaluation guidelines of T_{2323} can be defined in Table B.3.3.

Table B.3.3 scoring guidelines for data validity

Score / Attribute	1	3	5
Validity T_{2323}	$T_{2323} = 1$ if none of the validity criteria (illustrated above) is fulfilled	$T_{2323} = 3$ if the validity criteria (illustrated above) are partially fulfilled	$T_{2323} = 5$ if all of the validity criteria (illustrated above) are fulfilled

Accuracy measures how close the estimated or measured value is compared to the true value. Accuracy is determined by random and systematic errors in the measurements (Popek 2017). Since the data involved in nuclear PRA are mostly related to the number of failures or degradations and are usually collected digitally from different sources, systematic errors in the data are very small. This means that the accuracy of data is primarily determined by random errors. Since the error margin of the confidence interval is widely accepted as a good indicator of the random errors, it can be used as a measure of the data accuracy. Error factor may be defined based on the upper and lower bounds of confidence interval:

$$error\ factor = \sqrt{\frac{U_l}{L_l}}$$

where U_l and L_l are the upper and the lower bounds of confidence intervals. The accuracy of data is, then, scored based on the value of error factors, following the guidelines in Table B.3.4 scoring guidelines for data reliability

Table B.3.4 scoring guidelines for data validity

Score / Attribute	1	3	5
Accuracy T_{2325}	$T_{2325} = 1$ if the error factor is greater than 10	$T_{2325} = 3$ if the error factor is between 2-10	$T_{2325} = 5$ if the error factor is less or equal to 2

The rest of the "leaf" attributes of the reliability of data are evaluated following the guidelines in Table B.3.5.

Table B.3.5 scoring guidelines for data reliability

Score / Attribute	1	3	5
Consistency T_{2322}	$T_{2322} = 1$ if the data are not from the same type of power plant, or have different characteristics compared to the system under investigation, e.g., different component or model	$T_{2322} = 3$ if the data are from the same power plant with the same type of component and the same characteristics of the system under investigation but from different manufacturers	$T_{2322} = 5$ if the data are from the same power plant with the same type of components and the components have the same characteristics and the same manufacturer
Timeliness T_{2324}	$T_{2324} = 1$ if the data has never been updated	$T_{2324} = 3$ if the data has been updated a few years ago (10 years and more)	$T_{2324} = 5$ if the data are up-to-date and are updated routinely

Appendix VI :

Synthèse de thèse

Synthèse de thèse

L'objectif de l'évaluation des risques est de fournir un support d'informations pour la prise de décision [35], [36], [5], [3], [34]. Dans l'évaluation de risque, nous effectuons des mesures quantitatives et qualitatives du risque pour s'assurer qu'il reste dans la limite autorisée. L'évaluation quantitative du risque est effectuée par l'agrégation des risques multiples (MHRA), qui implique l'agrégation des indices de risque des contributeurs (de risque) pour arriver à une métrique de risque qui peut être comparée aux critères de sûreté pour aider à la prise de décision. D'un côté, en MHRA, les indices de risques des différents contributeurs peuvent avoir différents degrés de réalisme, qui résultent des différences dans leurs caractérisations, comme par exemple, leur incertitude, niveau de connaissance, conservatisme, etc. [19]. D'un autre côté, la pratique actuelle de la méthode MHRA consiste à effectuer une sommation arithmétique simple des indices de risque des différents contributeurs, sans considérer les aspects qui conduisent à la différence des degrés de réalisme [19]. La méthode MHRA doit donc considérer les différences d'incertitudes [19] et de degré de confiance dans les résultats (de l'évaluation de risque) qui sont pertinents pour soutenir la prise de décision [3].

Cette thèse de doctorat aborde le problème de l'agrégation de risques multiple (MHRA), qui vise à agréger les risques estimés pour différents contributeurs. La pratique actuelle de la MHRA est basée sur une sommation arithmétique simple des estimations de risques. Cependant, ces estimations sont obtenues à partir de modèles EPS (Estimation Probabiliste de risque) qui présentent des degrés de réalisme différents liés à différents niveaux de connaissances. En ne prenant pas en compte ces différences, le processus MHRA pourrait conduire à des résultats trompeurs pour la prise de décision (DM). Dans cette thèse, un cadre structuré est proposé afin d'évaluer le niveau de réalisme et de confiance dans les évaluations de risques et de l'intégrer dans le processus de MHRA. Ces travaux ont permis :

(i) Une identification des facteurs contribuant à la fiabilité de l'évaluation des risques. Leurs criticités sont analysées afin de comprendre leur influence sur l'estimation des risques ;

(ii) Un cadre hiérarchique intégré est développé pour évaluer la confiance et le réalisme de l'estimation de risque, sur la base des facteurs et des attributs identifiés en (i) ;

(iii) Une méthode basée sur un modèle réduit est proposée pour évaluer efficacement la fiabilité de l'évaluation des risques dans la pratique. Grâce à cette méthode, le nombre d'éléments pris en compte dans l'évaluation initiale des risques peut être limité ;

(iv) Une technique qui combine la théorie de Dempster-Shafer et le processus de hiérarchie analytique (DST-AHP) est appliquée au modèle développé. Cette technique permet d'évaluer le niveau de réalisme et confiance -dans l'analyse de risque- en utilisant une moyenne pondérée des attributs: la méthode AHP est utilisée pour calculer le poids des attributs et la méthode DST est utilisée pour tenir compte de l'incertitude subjective dans le jugement des experts dans l'évaluation des poids ;

(v) Une technique de MHRA est développée sur la base d'un modèle de moyenne bayésienne afin de surmonter les limites de la pratique actuelle de MHRA qui néglige le réalisme et confiance dans l'évaluation de chaque contributoire de risque ;

(vi) Le modèle développé est appliqué sur des cas réels de l'industrie des centrales.

Le modèle développé fournit un moyen systématique pour évaluer la fiabilité des résultats de l'évaluation des risques et pour les l'intégrer dans l'agrégation des risques afin de combler les lacunes de la MHRA conventionnelle. D'un point de vue pratique, l'approche prévoit également des procédures systématiques et pratiques pour faciliter son application sur des problématiques réelles et résoudre le problème de la subjectivité des jugements des experts. L'application du modèle développé sur des problématiques réelles démontre la faisabilité et le caractère raisonnable de l'approche, ouvrant la voie à son applicabilité pour aider la prise de décision basée sur les risques.