A practical approach for evaluating the strength of knowledge supporting risk assessment models

Tasneem Bani-Mustafa, Zhiguo Zeng, Enrico Zio, Dominique Vasseur

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 and Aven, 2009). These probabilistic indexes express the irreducible “aleatory uncertainty” in the related systems and processes (Helton and Burmaster, 1996), (Helton *et al.,* 2004), (Flage and 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 (Flage and Aven, 2009). Lack of knowledge will result in additional uncertainty in the PRA results, known as “epistemic uncertainty” (Helton and Burmaster, 1996), (Helton *et al.,* 2004), (Flage and 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).

However, the existing works on epistemic uncertainty quantification and propagation (for example, including but not limited to subjective probability, law of total expectations, 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 the law of total expectation, a probability distribution expressing the belief on different assumptions is introduced and then propagated. Compared to the uncertainty propagation, how to represent the epistemic uncertainty in the input parameters is less discussed in literature. With respect to this problem, assessing 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.*, 2017). 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, Goerlandt and Kujala, 2014), (Aven, 2012), (Aven and Ylönen, 2016), (Aven, 2013b), (Bjerga and 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 and 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 and Aven, 2012), (Aven, 2013b), (Montewka et al., 2014), (Goerlandt and Montewka, 2015), (Valdez Banda *et al.*, 2015), (Berner and Flage, 2016a), (Berner and Flage, 2016b), (Goerlandt and 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 and Aven, 2012), (Aven, 2013b), (Montewka *et al.,* 2014), (Goerlandt and Montewka, 2015), (Valdez Banda *et al.*, 2015), (Goerlandt and Reniers, 2016), (Berner and 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).

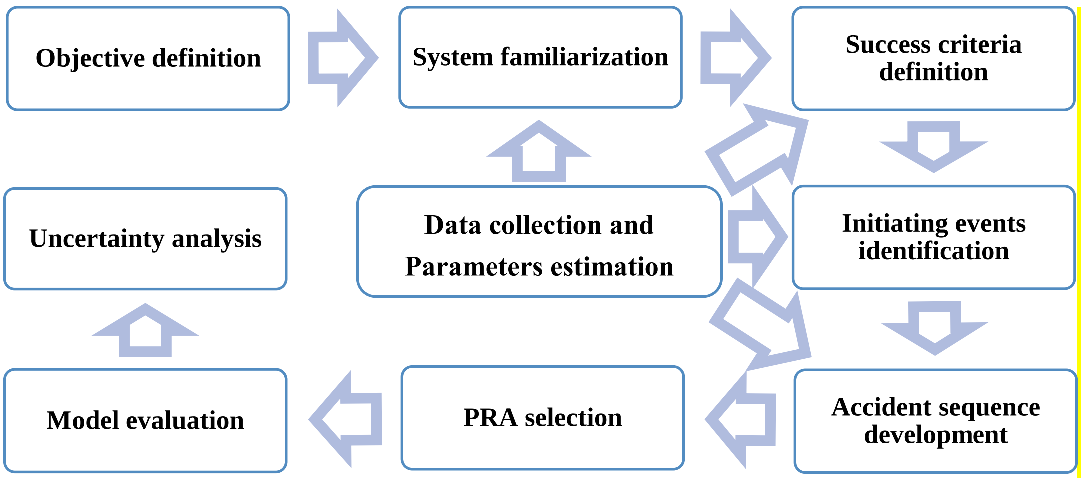

Figure 1 Typical PRA process flow

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	• The evolution sequence is known and well represented *(explicit and implicit knowledge in forms of phenomenological understanding)*

and clearly defined, and cover all the possible scenarios	• 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 (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 and 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 and 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 and Aven, 2009), (Goerlandt and 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. Note that any kind of input or conditions that are assumed and acknowledged to possibly deviate from reality are considered assumptions, e.g., input data that are assumed are considered a part of assumptions and not data.

1

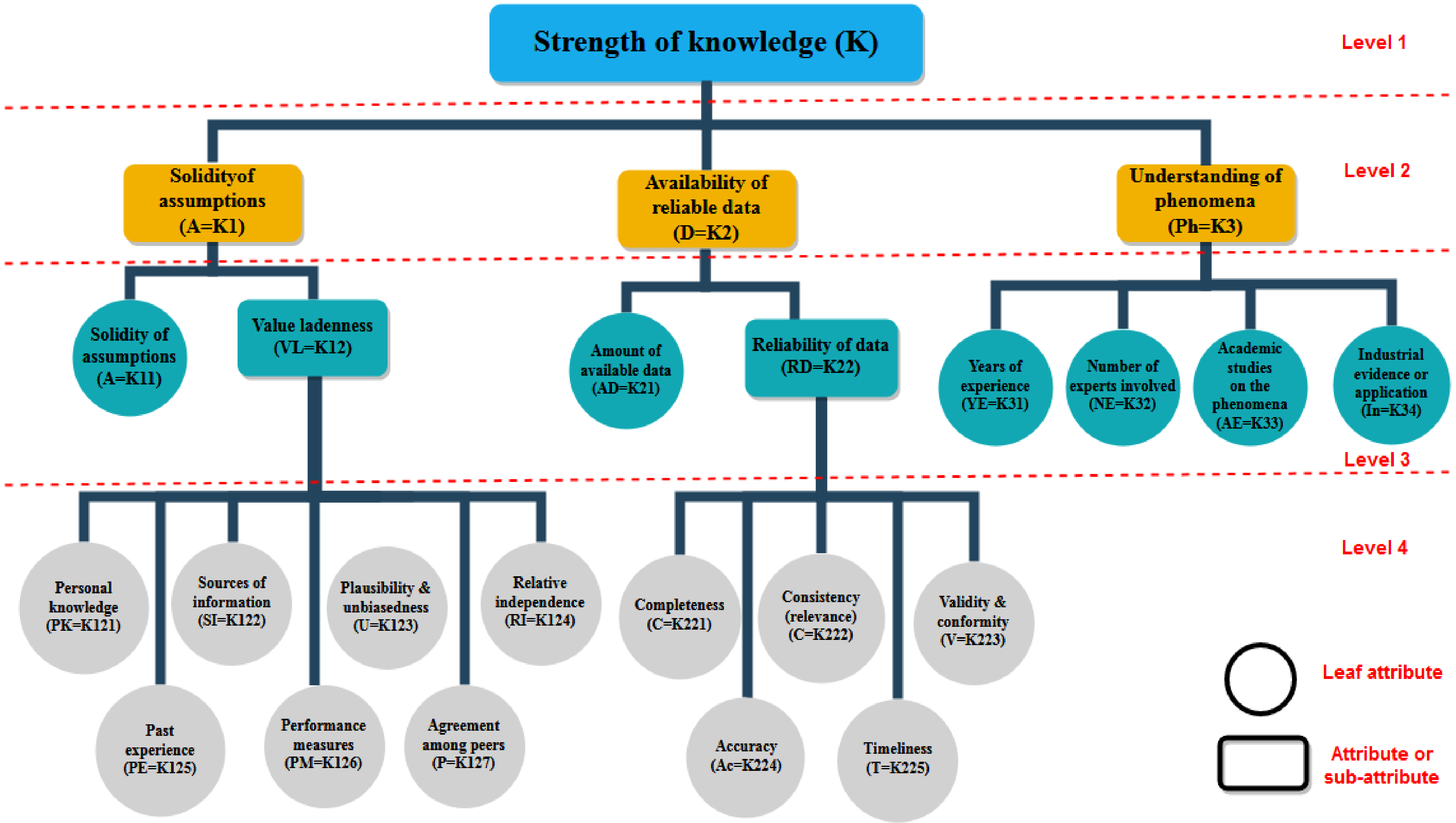

Figure 2 A hierarchical conceptual framework for knowledge assessment

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.

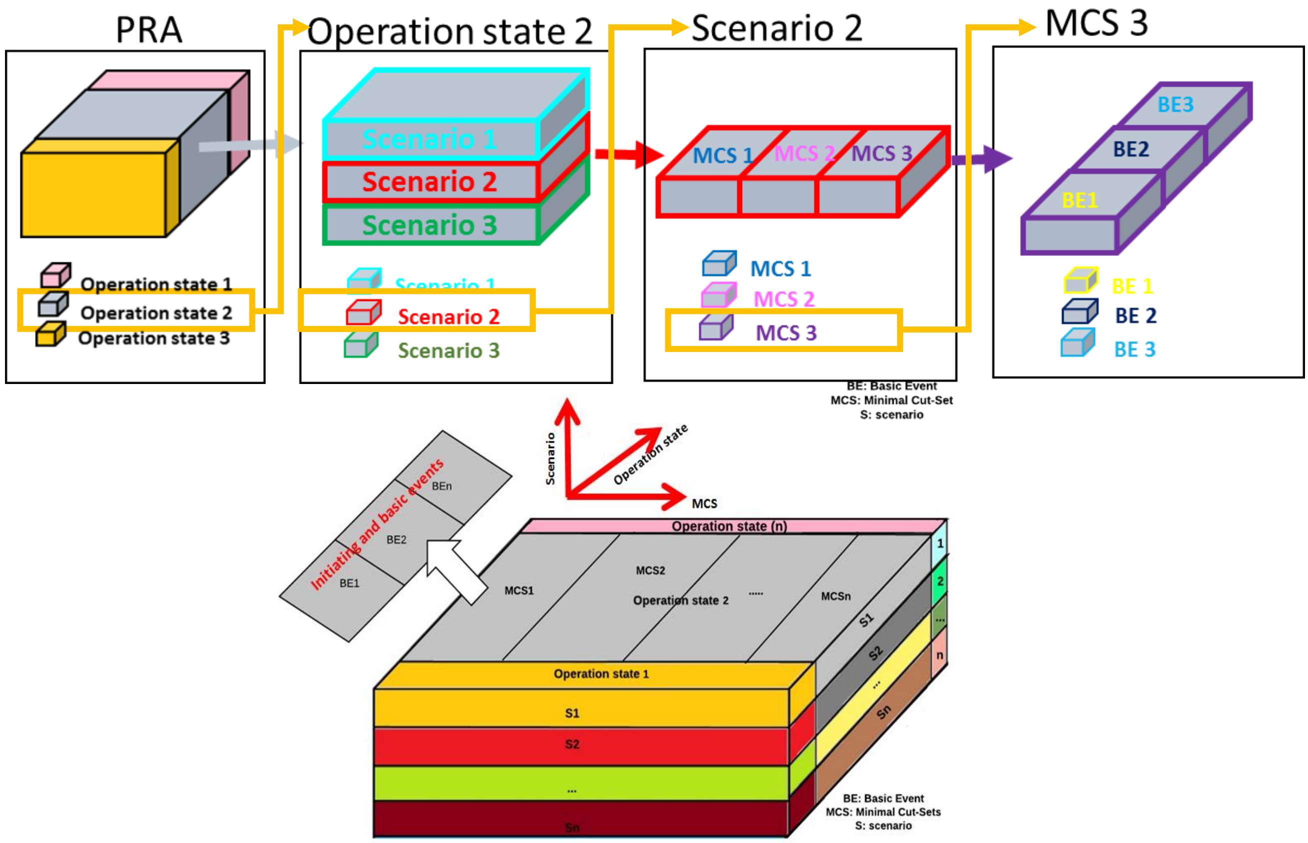

Figure 3 Atomic elements of a PRA model

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, 1998). 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.

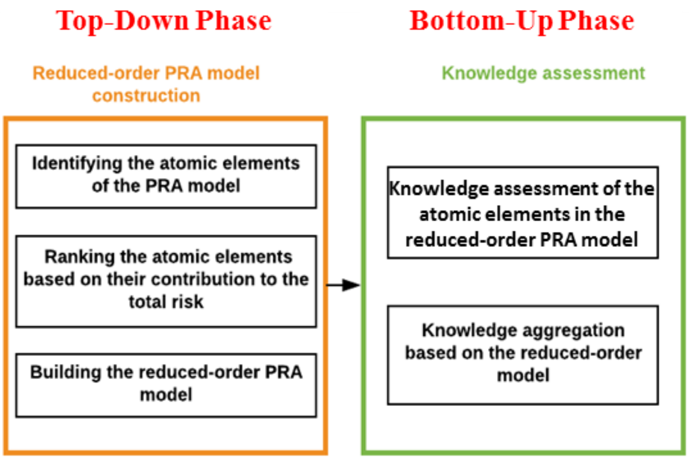

Figure 4 Procedural steps of the developed method

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*" (Hardy, 2010)). 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:

$$\frac{\sum_{i=1}^{n_{O,Red}} R_{O_i}}{R} \geq \alpha, \quad (4)$$

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 $R_{i,j}$, $1 \leq j \leq n_{S,i}$ by (2).
- Rank $R_{i,j}$ in descending order, $1 \leq j \leq n_{S,i}$.
- Find the minimal $n_{S,Red,i}$ so that:

$$\frac{\sum_{j=1}^{n_{S,Red,i}} R_{i,j}}{R_{O,i}} \geq \beta, \quad (5)$$

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 $i = 1,2, \dots, n_{O,Red}$.

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:

$$R_{i,j,k} = \prod_{q \in MCS_{i,j,k}} P_{BE,q}, \quad {}_{\substack{1 \leq i \leq n_{O,Red} \\ 1 \leq j \leq n_{S,Red,i} \\ 1 \leq k \leq n_{MCS,i,j}}}, \quad (6)$$

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

$$\frac{\sum_{k=1}^{n_{MCS,Red,i,j}} R_{i,j,k}}{R_{i,j}} \geq \gamma, \quad (7)$$

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):

$$R_{Red,l} = \sum_{k=1}^{n_{MCS,Red,l}} \prod_{q \in MCS_{l,k}} P_{BE,q}, \tag{9}$$

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:

$$K = \sum_{i=1}^{n_i} \sum_{j=1}^{n_{ij}} \sum_{k=1}^{n_{ijk}} W_i . W_{ij} . W_{ijk} . K_{ijk}, \tag{10}$$

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:

$$K = \sum_{k=1}^{n_{leaf}} W_{global,k} . K_{leaf,k}, \tag{11}$$

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:

$$W_{global,k} = \begin{cases} W_i . W_{ij}, & if K_{leaf,k} \text{ is in level 3} \\ W_i . W_{ij} . W_{ijk}, & if K_{leaf,k} \text{ is in level 4'} \end{cases} \tag{12}$$

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, $K_{leaf,k}$ 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 and 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 and 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 and Vargas, 2012), (Mu and 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 and Lamata, 2006), (Saaty and 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:

$$K_{BE} = \sum_{k=1}^{n_{leaf}} W_{global,k} \cdot K_{leaf,k} \tag{15}$$

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, 2009a). 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

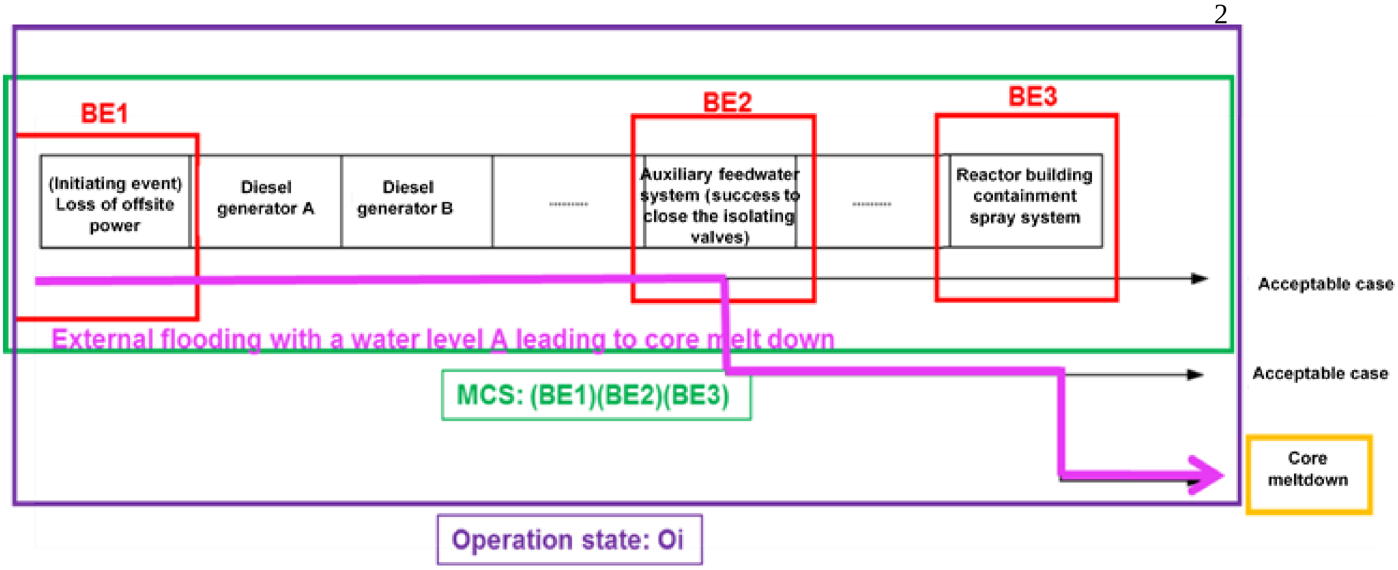

Figure 5 Illustration of a MCS in an individual reduced-order model

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, 2009b). 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.

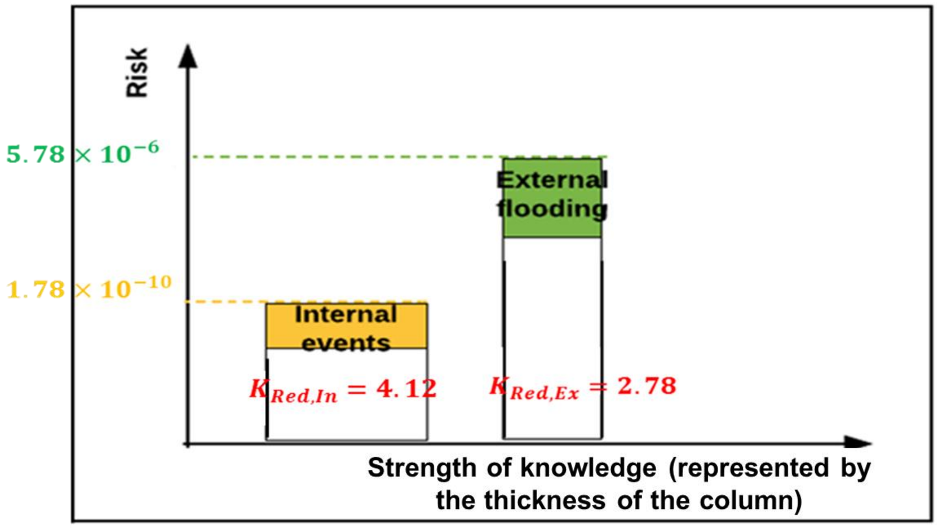

Figure 6 Representation of hazard groups levels of risk and SoK

5. 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 and 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 and Mosleh, 2008), (Droguett, 1999). Furthermore, the use of AHP method does not allow considering the interdependencies that might exist between some attributes. 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. and 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. doi: https://doi.org/10.1016/j.ress.2016.06.011.
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. and 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. doi: 10.1142/S0218488506004114.
4. Apostolakis, G. (1990) 'The concept of probability in safety assessments of technological systems', *Science*. American Association for the Advancement of Science, 250(4986), pp. 1359–1364.
5. Askeland, T., Flage, R. and Aven, T. (2017) 'Moving beyond probabilities ??? Strength of knowledge characterisations applied to security', *Reliability Engineering and System Safety*. Elsevier, 159(October 2016), pp. 196–205. doi: 10.1016/j.ress.2016.10.035.
6. Aven, T. (2012) 'The risk concept historical and recent development trends', *Reliability Engineering & System Safety*, 99, pp. 33–44. doi: https://doi.org/10.1016/j.ress.2011.11.006.
7. 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. doi: 10.1016/j.ress.2012.09.014.
8. Aven, T. (2013b) 'Practical implications of the new risk perspectives', *Reliability Engineering and System Safety*, 115, pp. 136–145. doi: 10.1016/j.ress.2013.02.020.
9. Aven, T. (2017a) 'How some types of risk assessments can support resilience analysis and management',

Reliability Engineering & System Safety, 167, pp. 536–543. doi: https://doi.org/10.1016/j.ress.2017.07.005.

10. 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. doi: https://doi.org/10.1016/j.ress.2017.05.006.
11. Aven, T. and 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. doi: https://doi.org/10.1016/j.ress.2013.07.005.
12. Aven, T. and Ylönen, M. (2016) 'Safety regulations: Implications of the new risk perspectives', *Reliability Engineering & System Safety*, 149, pp. 164–171. doi: https://doi.org/10.1016/j.ress.2016.01.007.
13. Bani-Mustafa, T., Pedroni, N., Zio, E., Dominique, V. and Francois, B. (2017) '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)*. ANS, pp. 314–323.
14. Bergdahl, M., Ehling, M., Elvers, E. and Földesi, E. (2007) *Handbook on Data Quality Assessment Methods and Tools*.
15. Berner, C. and Flage, R. (2016a) 'Comparing and integrating the NUSAP notational scheme with an uncertainty based risk perspective', *Reliability Engineering & System Safety*. Elsevier, 156, pp. 185–194.
16. Berner, C. and Flage, R. (2016b) 'Strengthening quantitative risk assessments by systematic treatment of uncertain assumptions', *Reliability Engineering and System Safety*, 151, pp. 46–59. doi: 10.1016/j.ress.2015.10.009.
17. Bjerga, T. and 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. doi: https://doi.org/10.1016/j.ress.2014.10.013.
18. Boone, I., Van der Stede, Y., Dewulf, J., Messens, W., Aerts, M., Daube, G. and Mintiens, K. (2010) 'NUSAP: a method to evaluate the quality of assumptions in quantitative microbial risk assessment', *Journal of Risk Research*. Taylor & Francis, 13(3), pp. 337–352. doi: 10.1080/13669870903564574.
19. Chen, H., Hailey, D., Wang, N. and Yu, P. (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. doi: 10.3390/ijerph110505170.
20. Coyle, G. (2004) *The analytic hierarchy process (AHP)*, *Practical strategy: Structured tools and techniques*. New York.

21. DAMA (2013) *The six primary dimensions for data quality assessment: defining data quality dimensions*. United Kingdom.
22. Davies, M. (2001) ‘Knowledge (explicit and implicit): philosophical aspects’, *International Encyclopedia of the Social and Behavioral Sciences*, 12, pp. 8126–8132.
23. Droguett, E. L. (1999) *Methodology for the treatment of model uncertainty*, *Proceedings of the 5th International Conference on Probabilistic Safety Assessment and Management (PSAM-5)*. University of Maryland at College Park.
24. Droguett, E. L. and Mosleh, A. (2008) ‘Bayesian methodology for model uncertainty using model performance data’, *Risk Analysis*. Wiley Online Library, 28(5), pp. 1457–1476.
25. 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.
26. EPRI (2015) *An Approach to Risk Aggregation for RisK-Informed Decision-Making*. Palo Alto, California.
27. Flage, R. and Aven, T. (2009) ‘Expressing and communicating uncertainty in relation to quantitative risk analysis’, *Reliability: Theory & Applications*. Интернет-сообщество Gnedenko Forum, 4(2–1 (13)).
28. Goerlandt, F. and 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. doi: 10.1201/b17399-230.
29. Goerlandt, F. and Montewka, J. (2015) ‘Maritime transportation risk analysis: Review and analysis in light of some foundational issues’, *Reliability Engineering & System Safety*, 138, pp. 115–134. doi: https://doi.org/10.1016/j.ress.2015.01.025.
30. Goerlandt, F. and Reniers, G. (2016) ‘On the assessment of uncertainty in risk diagrams’, *Safety Science*, 84, pp. 67–77. doi: https://doi.org/10.1016/j.ssci.2015.12.001.
31. Haouzi, H. El, Charpentier, P., Haouzi, H. El, Charpentier, P. and Adaptive, T. (2013) ‘Toward Adaptive Modelling & Simulation for IMS : The Adaptive Capability Maturity Model and Future Challenges’, in *IFAC Proceedings Volumes*, pp. 174–179.
32. Hardy, M. (2010) ‘Pareto’s law’, *The Mathematical Intelligencer*. Springer, 32(3), pp. 38–43.
33. Helton, J. C. and Burmaster, D. E. (1996) ‘Guest editorial: treatment of aleatory and epistemic uncertainty in performance assessments for complex systems’, *Reliability Engineering & System Safety*. Elsevier Science Publishing Company, Inc., 54(2–3), pp. 91–94.

34. Helton, J. C., Johnson, J. D. and Oberkampf, W. L. (2004) ‘An exploration of alternative approaches to the representation of uncertainty in model predictions’, *Reliability Engineering & System Safety*. Elsevier, 85(1–3), pp. 39–71.
35. Hey, J. (2004) *The data, information, knowledge, wisdom chain: the metaphorical link. Intergovernmental Oceanographic Commission., 2004*. Available at: http://www.dataschemata.com/uploads/7/4/8/7/7487334/dikwchain.pdf.
36. IAEA (1991) *Data Collection and Record Keeping for the Management of Nuclear Power Plant Ageing*. Edited by IAEA.
37. IAEA (2003) *External Events Excluding Earthquakes in the Design of Nuclear Power Plants*.
38. IAEA (2009a) *Deterministic Safety Analysis for Nuclear Power Plants*.
39. IAEA (2009b) *Meteorological and Hydrological Hazards in Site Evaluation for Nuclear Installations*.
40. IAEA (2010) *Development and Application of Level 1 Probabilistic Safety Assessment for Nuclear Power Plants*.
41. De Jong, A., Wardekker, J. A. and Van der Sluijs, J. P. (2012) ‘Assumptions in quantitative analyses of health risks of overhead power lines’, *Environmental science & policy*. Elsevier, 16, pp. 114–121.
42. Kelp, C. (2015) ‘Understanding phenomena’, *Synthese*. Springer, 192(12), pp. 3799–3816.
43. Khorsandi, J. and Aven, T. (2017) ‘Incorporating assumption deviation risk in quantitative risk assessments: A semi-quantitative approach’, *Reliability Engineering & System Safety*, 163, pp. 22–32. doi: https://doi.org/10.1016/j.ress.2017.01.018.
44. Kidwell, J. J., Vander Linde, K. and Johnson, S. L. (2000) ‘Applying corporate knowledge management practices in higher education.’, *Educause quarterly*. ERIC, 23(4), pp. 28–33.
45. Kloprogge, P., Van der Sluijs, J. P. and Petersen, A. C. (2011) ‘A method for the analysis of assumptions in model-based environmental assessments’, *Environmental Modelling and Software*. Elsevier Ltd, 26(3), pp. 289–301. doi: 10.1016/j.envsoft.2009.06.009.
46. Milazzo, M. F. and 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. doi: https://doi.org/10.1016/j.ress.2011.12.001.
47. Montewka, J., Goerlandt, F. and Kujala, P. (2014) ‘On a systematic perspective on risk for formal safety assessment (FSA)’, *Reliability Engineering & System Safety*, 127, pp. 77–85. doi: https://doi.org/10.1016/j.ress.2014.03.009.

48. Morgan, S. L. and Waring, C. G. (2004) *Guidance on Testing Data Reliability*.
49. Mu, E. and Pereyra-Rojas, M. (2017) 'Understanding the analytic Hierarchy process', in *Practical Decision Making*. Springer, pp. 7–22.
50. Nowakowski, T., Młyńczak, M., Jodejko-Pietruczuk, A. and Werbińska-Wojciechowska, S. (2014) 'Safety and reliability: Methodology and applications', in. CRC Press, pp. 437–444.
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 (1998) *Risk Spectrum, PSA Professional 1.20/Teory Manual*, *RELCON AB*.
54. Roth, D. J. (2009) 'Assessing the Reliability of Computer-Processed Data', (July).
55. Rowley, J. and Hartley, R. (2017) *Organizing knowledge: an introduction to managing access to information*. Routledge.
56. Saaty, T. L. (2008) 'Decision making with the analytic hierarchy process', *International Journal of Services Sciences*, 1(1), p. 83. doi: 10.1504/IJSSCI.2008.017590.
57. Saaty, T. L. (2013) 'Analytic hierarchy process', in *Encyclopedia of operations research and management science*. Springer, pp. 52–64.
58. Saaty, T. L. and 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. doi: 10.1016/j.mcm.2007.03.022.
59. Saaty, T. L. and Vargas, L. G. (2012) 'How to make a decision', in *Models, methods, concepts & applications of the analytic hierarchy process*. Springer Science & Business Media, pp. 1–20.
60. Saltelli, A., Guimarães Pereira, Â., Van der Sluijs, J. P. and Funtowicz, S. (2013) 'What do I make of your latinorum? Sensitivity auditing of mathematical modelling', *International Journal of Foresight and Innovation Policy*. Inderscience Publishers Ltd, 9(2-3–4), pp. 213–234.
61. Van Der Sluijs, J. P., Craye, M., Funtowicz, S., Kloprogge, P., Ravetz, J. and Risbey, J. (2005) 'Combining Quantitative and Qualitative Measures of Uncertainty in Model-Based Environmental Assessment: The NUSAP System', *Risk Analysis*. Wiley Online Library, 25(2), pp. 481–492. doi: 10.1111/j.1539-6924.2005.00604.x.
62. Stamatelatos, M., Dezfuli, H., Apostolakis, G., Everline, C., Guarro, S., Mathias, D., Mosleh, A., Paulos, T., Riha, D. and Smith, C. (2011) *Probabilistic risk assessment procedures guide for NASA managers and practitioners*.
63. Valdez Banda, O. A., Goerlandt, F., Montewka, J. and Kujala, P. (2015) 'A risk analysis of winter navigation in Finnish sea areas', *Accident Analysis & Prevention*, 79, pp. 100–116. doi:

https://doi.org/10.1016/j.aap.2015.03.024.

64. 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. doi: 10.1016/0951-8320(96)00060-9.
65. Zio, E. (2007) *An introduction to the basics of reliability and risk analysis. Series in Quality*, *Reliability and Engineering Statistics*.

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.

$$\text{YoE} = \text{length of the data collection period (in years)} \times \text{sample size of the data}$$

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 (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)