A HIERARCHICAL TREE-BASED DECISION MAKING APPROACH FOR ASSESSING THE TRUSTWORTHINESS OF RISK ASSESSMENT M....

Tasneem Bani-Mustafa
CentraleSupélec

Nicola Pedroni
Politecnico di Torino

Enrico Zio
Politecnico di Milano

FRANCOIS BEAUDOUIN
Électricité de France (EDF)

A HIERARCHICAL TREE-BASED DECISION MAKING APPROACH FOR ASSESSING THE TRUSTWORTHINESS OF RISK ASSESSMENT MODELS

Tasneem Bani-Mustafa1[(1)], Nicola Pedroni[(1)], and Enrico Zio[(1), (2)], Dominique Vasseur[(3)] & Francois Beaudouin[(4)]
[(1)] *Chair on System Science and the Energetic Challenge, EDF Foundation*
Laboratoire Genie Industriel, CentraleSupélec/Université Paris-Saclay, Grande voie des Vignes, 92290 Chatenay-Malabry, France *tasneem-adeeb.bani-mustafa@centralesupelec.fr*
[(2)] *Energy Department, Politecnico di Milano, Via Ponzio 34/3, Milan, 20133, Italy*
[(3)] *EDF R&D, Management des Risques Industriels*
EDF Lab Paris Saclay - 7 Bd Gaspard Monge, 91120 Palaiseau, France
[(4)] *EDF R&D, Management des Risques Industriels*
EDF Lab Chatou - 6 Quai Watier 78401 Chatou

Abstract: Risk assessment models are conceptual constructs (translated into mathematical forms), built on a set of assumptions (hypotheses) made on the available knowledge. In this sense, the risk assessment outcomes are conditional on the available knowledge.

Risk assessment provides informative support to decision making (DM), and assurance must be provided to guarantee that the results are credible and trustworthy for the DM purposes, for which they are employed.

The present paper proposes a four-levels, top-down, hierarchical tree to identify the main attributes and criteria that affect the level of trustworthiness of models used in probabilistic risk assessment. Based on this hierarchical decomposition, a bottom up, quantitative approach is employed for the assessment of model trustworthiness, using tangible information and data available at the basic "leaf" sub-attributes level. The analytical hierarchical process (AHP) is adopted for evaluating and aggregating the sub-attributes.

The approach is shown by application to a case study concerning the estimation of failure probability of the Residual Heat Removal (RHR) system of a nuclear power plant (NPP). The trustworthiness of two models of different complexity is evaluated: a Fault Tree (FT) and a Multi-States Physics-based Model (MSPM).

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

I. INTRODUCTION

In general terms, risk describes the future *consequences* (usually seen in negative, undesirable terms with respect to the planned objectives) potentially arising from the operation of a system or an activity, and the associated *uncertainty* [1]. Risk should be quantitatively assessed in order to be possibly compare it with predefined safety criteria, for aiding risk-informed decision-making processes [2].

In recent times we have witnessed a vivid discussion on the fundamental concept of "risk" and related foundational issues regarding its assessment. From a general perspective, it is understood that the outcomes of risk assessments (i.e., the undesirable events/scenarios, consequences and the description of uncertainty about these) are conditioned on the *background knowledge* and *information* available on the system and/or process under analysis, including assumptions and presuppositions, phenomenological understanding, historical system performance data and expert judgments [3]; [4]; [5]; [6].

Then, the risk assessment outcomes may have a more or less solid foundation, depending on the validity of the assessment and hypotheses made, and quality and quantity of data and extents used: poor models, lack of data or simplistic assumptions are examples of potential sources of (model) uncertainty "hidden in the background knowledge" of a risk assessment [6].

As well known in practice the modeling of a system or process needs to balance between two conflicting concerns: (i) *accurate representation* of the phenomena and mechanisms in the system or process and (ii) definition of the proper *level of detail* of the description of the phenomena and mechanisms, so as to allow the timely and efficient use of the model. Differences between the real world quantities and the model outputs inevitably arise from the conflict of these two concerns [7]; [8].

Since (i) the importance placed on modeling and simulation is increasingly high within safety-critical system engineering contexts and (ii) the fundamental value of a risk assessment lies in providing informative support to (high-consequence) decision making [9]; [2], the *confidence* that can be put in the accuracy, representativeness and completeness of the models is fundamental and a satisfactory level of assurance must be provided that the results obtained from such models are *credible* and *trustworthy* for the decision-making purposes for which they are employed [4]; [8]; [10].

The objective of the present paper is to propose a four levels, top-down, hierarchical tree-based decision-making approach to assess the trustworthiness of models used in risk assessment. The level of trustworthiness is divided into two attributes (level 2), four sub-attributes (level 3), and seven basic "leaf" sub-attributes (level 4). At the bottom of the structure, we place the alternative models for which we want to assess the trustworthiness and credibility. On the basis of this hierarchical decomposition the level of trustworthiness is then calculated by resorting to a bottom-up, quantitative approach. The basic "leaf" attributes represent "tangible" attributes that can be directly and quantitatively evaluated using data and information available (e.g., past knowledge, experts, judgment, historical records, etc.). Finally, the Analytical Hierarchical Process (AHP) is employed for evaluating and aggregating the sub attributes.

The proposed approach has been applied to assess the trustworthiness of two models (of different complexity and level of detail) of a Residual Heat Removal (RHR) System of the CPY900 Nuclear Power Plant (NPP) [11]: the two models are used to estimate the failure probability of the safety system of interest. The first model is based on a classical Boolean logic-based Fault Tree (FT). The components' failure rates are based on field data and/or expert judgment. The model does not consider possible dependencies between the states of degradation of different components (e.g., a valve and a pump) nor the interaction between physical and environmental parameters, and the mechanisms of components' degradation [11]. On the other hand, the second approach is based on a Multi-States Physics-based Model (MSPM). The analysis takes into account multiple time-dependent component's degradation states, the effect of physical and environmental parameters on the mechanisms of degradation and the dependency existing between the degradation of components [12]; [13].

A review of approaches for assessing the trustworthiness and credibility of a model is presented in Section II. In Section III, we present a hierarchical tree-based decision making approach for assessing model trustworthiness. In Section IV, we apply the proposed framework to the case study of the RHR system of a NPP. Finally, in Section V, we discuss the results and close the paper with some conclusions.

II. TRUSTWORTHINESS AND CREDIBILITY OF RISK ASSESSMENT MODELS

In this section, we review some approaches proposed in the literature to assess the trustworthiness and credibility of mathematical models. This has been, and still is, an issue of great significance in the nuclear industry, for the need of assuring the technical adequacy and trustworthiness of Probabilistic Risk Assessment (PRA) models. Indeed, it is an issue that has been at the forefront of using PRA for decision making for many years. In the Regulatory Guide RG 1.174 of the United States Nuclear Regulatory Commission (NRC), PRA analysis trustworthiness and appropriateness (how well the risk is assessed) in the context of decision making are addressed with respect to the scope, level of detail, technical adequacy and plant representation [17]. The adequacy of the actual modeling and the reasonableness of the assumptions and approximations made are considered, and the full comprehension and inclusion of PRA elements are emphasized together with the need of addressing the impact of uncertainty [17]. Following up, RG 1.200 provides concrete guidelines on technical elements for a technically acceptable PRA, its peer review program documentation etc., to apply RG 1.174 for evaluating whether a PRA is sufficiently adequate and trustworthy to support decision making[18]. More recently, EPRI 3002003116 emphasizes the importance of evaluating the maturity and trustworthiness of risk assessments regarding different hazard groups of different natures and proposes an approach to overcome these difficulties within the RG.1174 context [19]. While these efforts do not use a hierarchical quantification approach; they do attempt to establish criteria and processes by which technical adequacy, mature, peer review, and other aspects are used to build confidence in the use of PRA models for regulatory risk-informed decision making use.

In the literature, the trustworthiness of a method or a process is often measured in terms of its maturity. The concept of model maturity goes back to the 1970s: at the time, it was used to assess the maturity of a function of an information system [14]; [15]; [16]. Later, the Software Engineering Institute (SEI) has developed a framework (the so-called Capability Maturity Model (CMM)) to assess the maturity of a software development process, in view of its quality, reliability and trustworthiness. Recently, the CMM model has been extended into the Prediction Capability Maturity Model (PCMM) for evaluating and assessing the maturity of modeling and simulation efforts [14]. Other examples of maturity assessment approaches have been developed in different domains, such as master data maturity assessment, enterprise risk management and hospital information system [16]. Finally, a hierarchical framework based on the analytical hierarchical process (AHP) has been developed to assess the maturity and prediction capability of a prognostic method for maintenance decision making purposes [16].

On the other hand, qualitative and semi-quantitative approaches have been proposed for evaluating the "strength-of-knowledge" in risk assessment models. In [3] a "crude" qualitative, direct grading of the strength-of-knowledge that supports risk assessment based on (mathematical) models is introduced. Actually, the authors try to classify the strength-of-knowledge to {minor, moderate, significant} with respect to the following elements [3]; [6]; [20]; [5]: (i) phenomenological understanding of the problem and availability of precise and well-understood predicting models for the physical phenomena of interest; (ii) availability of reliable data; (iii) reasonability of assumptions made (i.e., the assumptions do not exhibit large simplifications); (iv) agreement (consensus) among experts (i.e., low value ladenness). The strength-of-knowledge is, then, classified according to the following criteria [3]; [5]; [6]; [20]: (1) if none of the previously mentioned components is met, then the knowledge is "weak"; (2) if the "requirements" are partially met, then the strength-of-knowledge is considered "intermediate"; (3) if all "requirements" are met, then the knowledge is considered "strong".

In [20], a more detailed, semi-quantitative approach (namely the "assumption deviation risk") has been introduced. This approach is based on the identification of all the main assumptions on which the analysis is based. Then, the assumptions are converted into uncertainty factors and a rough evaluation of the deviation from the conditions defined by the assumptions is carried out. Finally, a score is assigned to each deviation that reflects the risk related to the deviation and its implications on the occurrence of given events and their consequences.

In [6], the authors embrace, apply, test and adjust the perspectives of [3] and [20] to develop a general and systematic framework for treating (uncertain) assumptions in risk assessment models. Also, this methodology for assessing the importance of assumptions is based on evaluating the basic components of the risk description mentioned above and previously developed and adopted by [20]. The evaluation places an assumption in one of six "settings", each providing guidelines to characterize the corresponding uncertainty. In practice, these guidelines are based on the precept that the effort that should be exerted for characterizing the uncertainty associated to an assumption and the effect on the related potential deviations, should increase with the importance and the criticality of the assumption. Finally, also in [8] the effect and importance of "structural" assumptions, approximations and simplifications on risk assessment model outputs [21] is studied by means of different approaches, including subjective and imprecise probabilities, and semi-quantitative scores (reflecting the degree of uncertainty associated to an assumption and the sensitivity of the model output to the assumption). The analysis serves as an input to the decision makers to understand which assumptions are unacceptable and need "remodeling".

III. HIERARCHICAL TREE-BASED DECISION MAKING APPROACH FOR ASSESSING THE TRUSTWORTHINESS OF RISK ASSESSMENT MODELS

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

III.A. Hierarchical tree for model trustworthiness: extraction and decomposition

Many factors affect the trustworthiness and credibility of models outcomes. Although they might sensibly vary depending on the problem at hand, some key factors can be summarized as follows: (i) phenomenological understanding of the problem; (ii) availability of reliable data; (iii) reasonability of the assumptions made; (iv) agreement among the experts; (v)

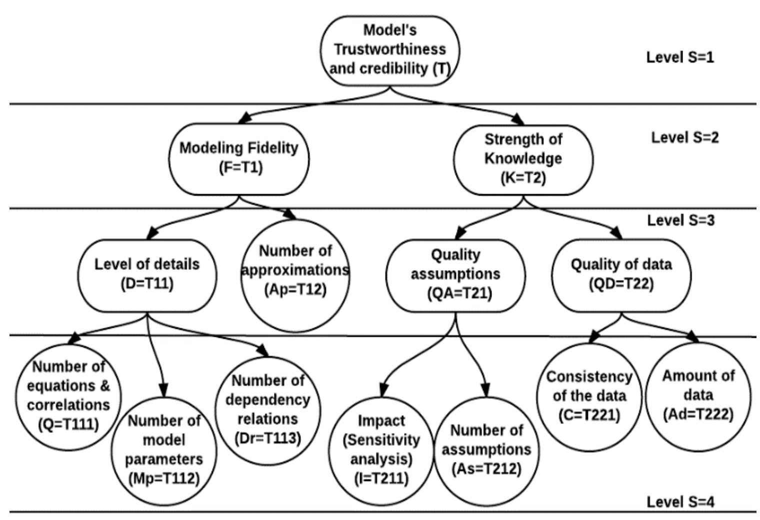

Figure 1 A hierarchical tree-based "decomposition" of the level of trustworthiness and credibility of a model

level of detail in the description of the phenomena and process of interest; (vi) accuracy and precision in the estimation of the values of the model's parameter; (vii) level of conservatism; (viii) amount of uncertainty, see e.g., [3]; [6]; [4]; [5]; [22]; [8]; [16]; [14]; [19]; [9]. Some of these attributes (criteria) are not "tangible" and cannot be "measured" directly: as a consequence, other sub-attributes should be identified, which can be easily measured and/or qualitatively evaluated. To this aim, we propose a method for model trustworthiness characterization and decomposition, which is based on an analytical hierarchical tree, such as the one in Figure 1.

The model trustworthiness, represented by T (Level 1), is characterized by two attributes: modeling fidelity, represented by $F = T_1$ and strength-of-knowledge, represented by $K = T_2$ (Level 2). The modeling fidelity ($F = T_1$), measures the adequacy of the model representation of the phenomenon and the level of detail adopted in the model description. The strength-of-knowledge ($K = T_2$) measures how "solid" the assumptions, data and information (on which the model relies) are. These two attributes are in turn decomposed into sub-attributes (Level 3). In particular, the modeling fidelity $F = T_1$ is defined by level of detail represented by $D = T_{11}$ (Level 3), and by the number of approximations $Ap = T_{12}$, whereas the strength-of-knowledge $K = T_2$ is defined by the quality of assumptions represented by $QA = T_{21}$ and by quality of data $QD = T_{22}$. Note that the number of approximations $Ap = T_{12}$ is considered as a basic attribute, since it can be measured directly and, thus, it is not broken further into other attributes. The other three attributes of Level 3 are instead broken down into more basic "leaf" attributes that can be measured directly by "inspection" of the model whose trustworthiness we want to assess. In particular, the level of detail $\mathrm{D} = \mathrm{T}_{11}$ is characterized in terms of number of equations and correlations, namely $\mathrm{Q} = \mathrm{T}_{111}$, number of model parameters, namely $\mathrm{Mp} = \mathrm{T}_{112}$, and number of dependency relations, namely $\mathrm{Dr} = \mathrm{T}_{113}$. The overall quality of the assumptions $\mathrm{QA} = \mathrm{T}_{21}$ is measured by the number of the assumptions made, $\mathrm{As} = \mathrm{T}_{212}$, and by their impact $\mathrm{I} = \mathrm{T}_{212}$ (which can be assessed, e.g., by sensitivity analysis). Finally, the quality of the data $\mathrm{QD} = \mathrm{T}_{22}$ is described in terms of the amount of data available, namely $\mathrm{Ad} = \mathrm{T}_{221}$, and by the consistency of the data itself, namely $\mathrm{C} = \mathrm{T}_{222}$.

III.B. Analytical hierarchical process (AHP) for model trustworthiness quantification

Given the hierarchical tree in Figure 1, the assessment of model trustworthiness is carried out within a multiple criterion decision analysis (MCDA) framework [23]; [24]. In this setting, we suppose in all generality that the system, components, process or phenomena of interest for the risk assessment can be obtained by different mathematical models, $M1, M2, .. M_l, .., M_n$ of possibly different complexities and levels of detail. The task (i.e., the MCDA problem at hand) is to rank these alternative models with respect to their trustworthiness, in relation to the particular risk assessment problem of interest to support MCDA. In the present paper, the Analytical Hierarchical Process (AHP) is adopted for this [16].

In the AHP the top goal of the evaluation, i.e., the considered decision problem (in this case, the selection of the model with the highest trustworthiness), is placed at the first level of the hierarchy and it is usually decomposed into several sub-attributes distributed over different levels. Finally, the bottom level in the hierarchal tree-based AHP model contains the different alternatives to be ranked with respect to the top goal (i.e., in this case the level of trustworthiness) [25]; [16]. Through pairwise comparisons among the attributes of the same level, the alternative solutions, i.e., models, can be ranked with respect to the decision problem in the top level (i.e., the identification of the model with highest trustworthiness) [25]; [26]. The AHP model for model trustworthiness assessment is represented in Figure 1.

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

The intermediate priorities $W(T_i)$, $W(T_{ij})$ and $W(T_{ijk})$ are calculated using pairwise comparison matrices: in particular a pairwise comparison matrix is constructed for each attribute, sub-attribute and basic "leaf" sub-attribute. A comparison matrix is a $(n \times n)$ square matrix where *n* is the number of elements being compared. Attributes in each level are compared to each other with respect to their importance in describing their "parent" attribute in the upper level. For example, a (3×3) matrix is constructed to compare the basic sub-attributes $Q = T_{111}$, $Mp = T_{112}$ and $Dr = T_{113}$ (Level 4) with respect to their" parent" sub-attribute $D = T_{11}$ (Level 3). Typically, a scale of 1 to 9 is chosen to evaluate the "strength" of each criteria with respect to the other; for example, a scale used to carry out a qualitative comparison between two attributes A and B can be as follows [25]; [27]:

1: A and B are equally important;
3: A is moderately more important than B;
5: A is strongly more important than B;
7: A is very strongly more important than B;
9: A is extremely more important than B.

Values 2, 4, 6 and 8 can be used to facilitate the judgment in intermediate situations.

A pairwise comparison matrix is made for each group of attributes in the same level (say s) that falls under the same upper attribute in the upper level (s-1). The weight of each attribute is, then, determined by solving an eigenvector problem, where the normalized principal eigenvector provides the weighting vector (priorities and strengths). Instead, for the tangible, basic, "leaf" sub-attributes T_{ijk} for which a quantitative evaluation can be given, the inter-level weights (or parties) $W(M_l, T_{ijk})$ can be obtained as:

$$W\left(M_l, T_{ijk}\right)\frac{T_{M_l,T_{ijk}}}{\sum_{l=1}^{n} T_{M_l,T_{ijk}}}, \qquad (1)$$

where $T_{M_l,T_{ijk}}$ is the numerical value that the basic "leaf" sub-attribute $T_{T_{ijk}}$ takes with respect to model M_l (for example, for attributes $Q = T_{111}$ variable $T_{M_l,T_{111}}$ equals the number of equations and correlations continued in M_l). Finally, notice that if the basic "leaf" sub-attributes cannot be given a direct numerical evaluation, the 1-9 scaling system explained above can be also adopted to evaluate the $T_{M_l,T_{ijk}}$'s. Notice that weights obtained should be normalized to 1 as follows: $\sum_{i=1}^{n_T} W(T_i) = 1$, where n_T is the number of attributes under the "top" attribute T (i.e., model trustworthiness); $\sum_{j=1}^{n_{T_i}} W(T_{ij}) = 1$, where n_{T_i} is the number of sub-attributes under attribute T_i; $\sum_{k=1}^{n_{T_{ij}}} W(T_{ijk}) = 1$, where n_{T_i} is the number of basic "leaf" sub-attributes under sub-attribute T_{ij}; and finally $\sum_{l=1}^{n} W(M_l, T_{ij}) = 1$, where n is the number of models. After obtaining the weight for each criterion with respect to the corresponding upper level criteria, a "global" weighting for each criterion with respect to the top goal T can also be obtained by multiplying its weight by the weights of its upper parent elements in each level: for example, the "global" weight (or priority) of basic "leaf" sub-attribute T_{iJk} with respect to the "top" attribute (goal) T is given by $W(T_{iJk}).\ W(T_{iJ}).\ W(T_i) = W_{global}(T_{ijk})$. For example, the global weighting of the consistency of data with respect to level of adequacy is obtained by multiplying its weight by the weight of quality of data by the weight of strength of the knowledge. Finally, the final trustworthiness $T(M_l)$ of a model M_l is evaluated using a weighted average of the "leaf" attributes as indicated in eq. (2):

$$T(M_l) = \sum_{i=1}^{n_T}\sum_{j=1}^{n_{T_i}}\sum_{k=1}^{n_{T_{ij}}} W(T_i).W(T_{ij}).W(T_{ijk}).W\left(M_l, T_{ijk}\right)\frac{T_{M_l,T_{ijk}}}{\sum_{l=1}^{n} T_{M_l,T_{ijk}}}, \qquad (2)$$

Where n_T, n_{T_i}, and $n_{T_{ij}}$ have been defined above. Equivalently, the trustworthiness $T(M_l)$ can be expressed directly as a function of the "global" weights of the leaf attributes with respect to the top goal T:

$$T(M_l) = \sum_{i=1}^{n_T}\sum_{j=1}^{n_{T_i}}\sum_{k=1}^{n_{T_{ij}}} W_{global}\left(T_{ijk}\right)\frac{T_{M_l,T_{ijk}}}{\sum_{l=1}^{n} T_{M_l,T_{ijk}}}, \qquad (3)$$

IV. CASE STUDY

In this section, the hierarchical tree-based trustworthiness assessment approach is applied to a case study concerning the modeling of the residual heat removal (RHR) system of a nuclear power plant (NPP). In section IV.A, the system is described; in section IV.B, the characteristics of the two models used to represent the system (i.e. the Fault Tree-FT and the Multi-States Physics-BASED Model-MSPM) are presented in some detail; finally, in section IV.C, the proposed approach is applied to evaluate the trustworthiness of the two models.

IV.A. The system

The Residual Heat Removal (RHR) system of the Electricité de France (EDF) CPY900 reactor is taken as reference (REF). RHR is mainly used to remove the decay heat (residual power) from the reactor cooling system and fuel during and after the shutdown, as well as supplementing spent fuel pool cooling in the shutdown cooling mode [28]. The main components of the RHR system are: pumps, heat exchangers, diaphragms, and valves.

According to a study implemented by EDF, it was found that 24.2% of RHR system failures are due to pumps failure, 64.4% are due to valves failures, 10.6% are due to heat exchanger failures, while only 0.8% of RHR system failures are due to other components' failure [11].

IV.B. Models considered

Two models have been considered for evaluating the reliability (resp., the failure probability) of the RHR system: a Fault Tree (FT) model (Section IV.B.1) and a more detailed Multi-State Physics-based Model (MSPM) (Section IV.B.2).

IV.B.1. Fault Tree (FT) Model

The Andromeda software has been used by EDF for the analysis of the RHRs' components failure modes and criticalities (importance analysis) [11]; [12]. Actually, this analysis provides a logical framework for understanding the different possible ways in which the components and the system can fail.

The failure probabilities used in the FT analysis provided by EDF are based on field experience [11].

IV.B.2. Multi-State Physics-based Model (MSPM)

In engineering systems, most products and components wear and degrade over time due to operational and mechanical factors, as well as their interaction with the surrounding environment and with each other [13]. There are many models of degradation processes in the bibliography. Physics-based model (PBM) and multi-state model (MSM) are often used for the degradation processes of components/systems. Physics-based model aims to develop an integrated mechanistic description of the component/system life, consistent with the underlying degradation mechanisms (e.g. wear, stress corrosion, shocks, cracking, fatigue, etc.) by using physics knowledge and equations, while multi-state model can be built upon material science knowledge, degradation and/or failure data from historical collection or degradation tests, to describe the degradation processes in a discrete way [29]; [13].

In general, MSM is able to describe the evolution of degradation with time, when there is a range of states from "perfect functioning" to "complete failure". However, since the degradation process is influenced by many factors, there are difficulties in estimating the transition rates required for the analysis of the degradation process, especially for highly reliable components and systems [13]. Besides, it is difficult to define precisely the states and the transition rates between states in MSMs, due to the imprecise discretization of the degredation process and to data insufficiency [12]. On the other hand, for PBM, the parameteres might be unknown so they are usually left to experts judgment. Accordingly, a combination of the two models, namely, the Multi-State Physics-based Model (MSPM), has been proposed, in which the state transition rate estimates are based on physical models rather than operation data [30]. Then, the whole process of transition and degradation can be described comprehensively by MSPM [13].

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

Figure 2 illustrates this setting. Three states were considered for the pump, including the fully functioning state, a degradation state corresponding to external leakage, and the failure state: thus, three transition rates were taken into account. The breaker was modeled by a continuous-time homogeneous Markov chain, taking into account the perfectly function, and the fully failed states, and four types of failures were taken into account. Similarly a continuous-time homogeneous Markov chain analysis was applied for the analysis of the contactor and the motor, and four and two types of failures were taken into account for each respectively.

On the other hand, the valve is subject to thermal fatigue that causes cracks or propagation of manufacture defects which are described by physical models and the related physical variables, such as; the coefficient of thermal expansion of the material, the modulus of elasticity, the Poisson ratio of the material, the elastoplastic strain concentration factors, the number of alternating cycles, etc.. The crack initiation takes place when the amplitude of variation of the critical temperature ΔT_{lim} is surpassed, while the failure due to propagation of defects takes place when a specific number of cycles (operation demands) is exceeded. It should be noted that the total number of cycles executed over a period of time is calculated considering the degradation dependency of the valves on the degradation of the pump. In other words, when calculating the number of cycles executed by the

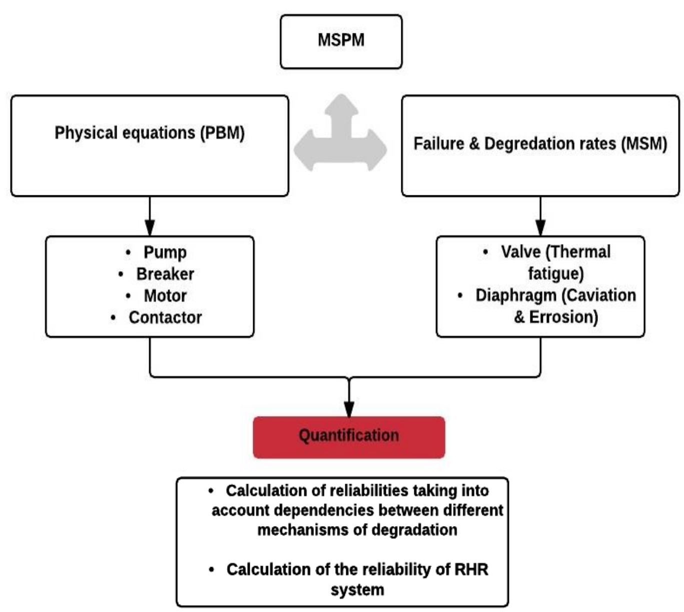

Figure 2 MSPM analysis: models of RHR components

valve, it is multiplied by a factor > 1 to consider the degradation dependency on the other components. Furthermore, the cavitation and the erosion are taken into account for analyzing the degradation and failure of the diaphragm. Different physical parameters are considered such as pressure, stress, dimension, and other material-based characteristics. A threshold value at which the failure takes place is taken into account. The results of MSPM and FT (using Andromeda software) are given in Table 1. The analysis shows similarities in the results in the first eight years. A gap between the two results starts to appear in the tenth year, showing a more rapid decline in the results obtained by MSPM.

Table 1 Values of reliability computed over time of Andromeda software (EDF)

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

IV.C. Quantitative evaluation

The analysis is carried out through two main steps: the first is an "upward" evaluation of the weight of each element in the hierarchy tree with respect the top goal (model trustworthiness); the second is a "downward" assessment of the model trustworthiness by means of a numerical evaluation of the basic "leaf" elements for both models FT and MSPM, as shown in Figure 3.

With respect to the weights evaluation, experts from EDF were asked to fill the pairwise comparison matrices, in order to evaluate the importance of each attribute (criteria). Experts were equally qualified and the inputs were averaged for simplicity. The attributes relative importance with respect to the parent attributes have been evaluated using the 1-9 scaling.

For the second step of "upward" calculation, two types of trustworthiness analysis have been implemented: one has been performed through a direct quantitative evaluation of the leaf attributes (e.g., the number of model parameters are counted, for each model); the second is based on a semi-quantitative evaluation of the leaf attributes carried out through comparing the two models to each other and to the state of the art, and then, assigning a relative score (1-9) for each leaf attribute.

Relying on the data and technical report [11] used by EDF to perform the risk assessment, the trustworthiness evaluation was performed for both **FT** and **MSPM** models, as illustrated in Tables 2 and 3.

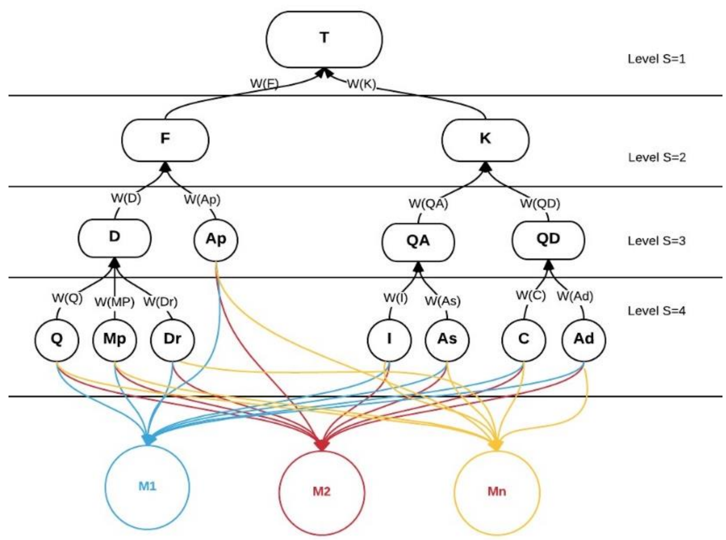

Figure 3 Hierarchical tree-based AHP model for the assessment of the trustworthiness of risk assessment models

The level of trustworthiness was found to be 4.6506 for FT (M1) and 5.8535 for MSPM (M2).

We applied the same method also to evaluate the models trustworthiness T using the direct quantification of the "leaf" attributes. The results are reported in Table 3 and Table 4 shows all results.

Table 2 Comparison between FT and MSPM trustworthiness

Parameter	Symbol	Level	Weight	Global weight	Fault Tree		MSPM	
					Score	Weighted score	Score	Weighted score
Model trustworthiness	**T**	**S1**	1	1	-	4.6506	-	5.8535
Modeling Fidelity	**F** (T_1)	**S2**	0.35	0.35	-	1.5124	-	2.3678
Number of approximations	**Ap** (T_{12})	**S3**	0.54	0.189	6	1.1340	7	1.323
Level of detail	**D** (T_{11})	**S3**	0.46	0.161	-	0.3784	-	1.0448
Number of equations and correlations	**Q** (T_{111})	**S4**	0.4638	0.0747	3	0.2240	8	0.5973
Number of model parameters	**Mp** (T_{112})	**S4**	0.2114	0.0340	3	0.1021	7	0.2383
Number of dependency relations	**Dr** (T_{113})	**S4**	0.3248	0.0523	1	0.0523	4	0.2092
Strength of Knowledge	**K** (T_2)	**S2**	0.65	0.65	-	3.1382	-	3.4857
Quality of data	**QD** (T_{22})	**S3**	0.51	0.3315	-	2.0553	-	2.2542
Amount of data	**Ad** (T_{221})	**S4**	0.6	0.1989	5	0.9945	8	1.5912
Consistency of data	**C** (T_{222})	**S4**	0.4	0.1326	8	1.0608	5	0.663
Quality assumptions	**QA** (T_{21})	**S3**	0.49	0.3185	-	1.0829	-	1.2315
Number of assumptions	**As** (T_{211})	**S4**	0.2	0.0637	5	0.3185	6	0.3822
Impact of the assumptions	**I** (T_{212})	**S4**	0.8	0.2548	3	0.7644	3.3333	0.8493

Table 3 Comparison between FT and MSPM level of trustworthiness (direct quantification)

Parameter	Symbol	Level	Weight	Global weight	Fault Tree		MSPM	
					Score	Weighted score	Score	Weighted score
Model trustworthiness	**T**	**S1**	1	1	-	57.15	-	111.7
Modeling Fidelity	**F** (T_1)	**S2**	0.35	0.35	-	0.374	-	1.3641
Number of approximations	**Ap** (T_{12})	**S3**	0.54	0.189	7	0.027	7	0.027
Level of detail	**D** (T_{11})	**S3**	0.46	0.161	-	0.347	-	1.3371
Number of equations and correlations	**Q** (T_{111})	**S4**	0.4638	0.0747	1	0.0747	9	0.672
Number of state rates and parameters	**Mp** (T_{112})	**S4**	0.2114	0.0340	8	0.2723	18	0.6128
Number of dependency relations	**Dr** (T_{113})	**S4**	0.3248	0.0523	0	0	1	0.0523
Strength of Knowledge	**K** (T_2)	**S2**	0.65	0.65	-	56.7772	-	110.33
Quality of data	**QD** (T_{22})	**S3**	0.51	0.3315	-	55.758	-	109.89
Amount of data	**Ad** (T_{221})	**S4**	0.6	0.1989	275	54.698	549.15	109.22
Consistency of data	**C** (T_{222})	**S4**	0.4	0.1326	8	1.0608	5	0.663
Quality assumptions	**QA** (T_{21})	**S3**	0.49	0.3185	-	1.0192	-	0.4463
Number of assumptions	**As** (T_{211})	**S4**	0.2	0.0637	4	0.2548	3	0.1911
Impact (Sensitivity analysis)	**I** (T_{212})	**S4**	0.8	0.2548	3	0.7644	3.3333	0.2552

Table 4 Summary of models trustworthiness

Normalized Trustworthiness	FT	MSPM
Scale 1-9 measures	0.4427	0.5573
Direct measures	0.3365	0.6635

V. DISCUSSION AND CONCLUSION

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

FT trustworthiness has been found to score 4.8205 out of 9, whereas MSPM has scored 5.8535 or 0.3365 and 0.6635, respectively, by direct measures of "leaf" attributes. The two results confirm the expectation that MSBM provides more trustworthy risk estimates than FT due to the fact that it takes into account components failure dependency relations and time dependency.

Of course, we do not claim that the trustworthiness approach is comprehensive and complete, as there exist other factors that affect the level of trustworthiness, which were not considered for simplicity. Also, further studies should be performed to define the scaling guidelines for attributes evaluation.

VI. REFERENCES

[1] INSAG, *A Framework for an Integrated Risk Informed Decision Making Process*, no. 25. Vienna: INTERNATIONAL ATOMIC ENERGY AGENCY, (2011).

[2] J. EISER *ET AL.*, "Risk interpretation and action: A conceptual framework for responses to natural hazards," *Int. J. Disaster Risk Reduct.*, vol. 1, no. 1, pp. 5–16, (2012).

[3] R. FLAGE AND T. AVEN, "Expressing and communicating uncertainty in relation to quantitative risk analysis," *Reliab. Risk Anal. Theory Appl.*, vol. 2, no. 13, pp. 9–18, (2009).

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

[5] H. VELAND AND T. AVEN, "Improving the risk assessments of critical operations to better reflect uncertainties and the unforeseen," *Saf. Sci.*, vol. 79, pp. 206–212, (2015).

[6] C. BERNER AND R. FLAGE, "Strengthening quantitative risk assessments by systematic treatment of uncertain assumptions," *Reliab. Eng. Syst. Saf.*, vol. 151, pp. 46–59, (2016).

[7] M. E. PATÉ-CORNELL, "Uncertainties in risk analysis: Six levels of treatment," *Reliab. Eng. Syst. Saf.*, vol. 54, no. 2, pp. 95–111, (1996).

[8] T. BJERGA, T. AVEN, AND E. ZIO, "An illustration of the use of an approach for treating model uncertainties in risk assessment," *Reliab. Eng. Syst. Saf.*, vol. 125, pp. 46–53, (2014).

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

[10] R. FLAGE AND T. AVEN, "Emerging risk – Conceptual definition and a relation to black swan type of events," *Reliab. Eng. Syst. Saf.*, vol. 144, no. August, pp. 61–67, (2015).

[11] EDF, "Programme de Base de Maintenance Préventive RRA palier 900 CPY," SAINT-DENIS CEDEX, (2001).

[12] Y.-H. LIN, Y.-F. LI, AND E. ZIO, "Fuzzy reliability assessment of systems with multiple-dependent competing degradation processes," *IEEE Trans. Fuzzy Syst.*, vol. 23, no. 5, pp. 1428–1438, (2015).

[13] F. DI MAIO, D. COLLI, E. ZIO, L. TAO, AND J. TONG, "A multi-state physics modeling approach for the reliability assessment of nuclear power plants piping systems," *Ann. Nucl. Energy*, vol. 80, pp. 151–165, (2015).

[14] W. L. OBERKAMPF, M. PILCH, AND T. G. TRUCANO, "Predictive capability maturity model for computational modeling and simulation," *cfwebprod.sandia.gov*, (2007).

[15] M. C. PAULK, B. CURTIS, M. B. CHRISSIS, AND C. V WEBER, "Capability Maturity Model for Software, Version 1.1," *Software, IEEE*, vol. 98, no. February, pp. 1–26, (1993).

[16] Z. ZENG, F. DI MAIO, E. ZIO, AND R. KANG, "A hierarchical decision-making framework for the

assessment of the prediction capability of prognostic methods," *Proc. Inst. Mech. Eng. Part O J. Risk Reliab.*, vol. 231, no. 1, pp. 36–52, (Dec. 2016).

[17] A. N. APPROACH *ET AL.*, "Assessment in risk-informed decisions on plant" (2011).

[18] A. N. APPROACH, F. O. R. DETERMINING, T. H. E. TECHNICAL, A. OF, P. RISK, AND R. F. O. R. R. ACTIVITIES, "An approach for determining the technical adequacy of probabilistic risk assessment results for risk-informed activities," no. March, (2009).

[19] EPRI, "An approach to risk aggregation for risk-informed decision-making" Palo Alto, California, (2015).

[20] T. AVEN, "Practical implications of the new risk perspectives," *Reliab. Eng. Syst. Saf.*, vol. 115, pp. 136–145, (2013).

[21] T. AVEN AND E. ZIO, "Model output uncertainty in risk assessment," *Int. J. Performability Eng.*, vol. 9, no. 5, pp. 475–486, (2013).

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

[23] L. XU AND J.-B. YANG, *Introduction to multi-criteria decision making and the evidential reasoning approach*. Manchester School of Management, (2001).

[24] E. TRIANTAPHYLLOU, B. SHU, S. N. SANCHEZ, AND T. RAY, "Multi-Criteria Decision Making : An Operations Research Approach," *Electronics*, vol. 15, pp. 175–186, (1998).

[25] T. L. SAATY, "Decision making with the analytic hierarchy process," *Int. J. Serv. Sci.*, vol. 1, no. 1, p. 83, (2008).

[26] E. ZIO, M. CANTARELLA, AND A. CAMMI, "The analytic hierarchy process as a systematic approach to the identification of important parameters for the reliability assessment of passive systems," *Nucl. Eng. Des.*, vol. 226, no. 3, pp. 311–336, (2003).

[27] E. ZIO, "On the use of the analytic hierarchy process in the aggregation of expert judgments," *Reliab. Eng. Syst. Saf.*, vol. 53, no. 2, pp. 127–138, (1996).

[28] NRC, *Reactor Coolant System and Connected Systems*. Washington: NRC, (2010).

[29] N. GORJIAN, L. MA, M. MITTINTY, P. YARLAGADDA, AND Y. SUN, "A review on degradation models in reliability analysis," in *Engineering Asset Lifecycle Management: Proceedings of the 4th World Congress on Engineering Asset Management (WCEAM 2009), 28-30 September 2009*, (2010), pp. 369–384.

[30] P. H. A. M. T. UNWIN SD, PP LOWRY, RF LAYTON, JR, "Multi-State Physics Models of Aging Passive Components in Probabilistic Risk Assessment," in *In International Topical Meeting on Probabilistic Safety Assessment and Analysis*, (2011), p. vol. 1, pp. 161–172.