APPORTS DE LA RECHERCHE

Uncertainty propagation and importance measure assessment

Case studies

Enrico Zio and Nicola Pedroni

n° 2013-12

THEME

Risk analysis

Les Cahiers de la sécurité industrielle

FONCSI
Fondation pour une culture de sécurité industrielle

The *Foundation for an Industrial Safety Culture* (FonCSI) is a french public-interest research foundation created in 2005. It aims to:

- undertake and fund research activities that contribute to improving safety in hazardous organizations (industrial firms of all sizes, in all industrial sectors);
- work towards better mutual understanding between high-risk industries and civil society, aiming for a durable compromise and an open debate that covers all the dimensions of risk;
- foster the acculturation of all stakeholders to the questions, tradeoffs and problems related to risk and safety.

In order to attain these objectives, the FonCSI works to bring together researchers from different scientific disciplines with other stakeholders concerned by industrial safety and the management of technological risk: companies, local government, trade unions, NGOs. We also attempt to build bridges between disciplines and to promote interaction and cross-pollination between engineering, sciences and the humanities.

The work presented in this document is the result of research funded by the FonCSI. The opinions presented are those of the authors.

FONCSI
Fondation pour une culture de sécurité industrielle

Foundation for an Industrial Safety Culture
A public-interest research foundation

http://www.FonCSI.org/

6 allée Émile Monso — BP 34038
31029 Toulouse cedex 4
France

Telephone: +33 534 32 32 00
Twitter: @LaFonCSI
Email: contact@FonCSI.org

Résumé

Titre	Études de cas en propagation d'incertitude et mesures d'importance
Mots-clefs	incertitude, probabilités, théorie des possibilités, analyse de risque, mesures d'importance
Auteurs	Enrico Zio et Nicola Pedroni
Date de publication	Décembre 2013

Les auteurs analysent l'impact de différentes représentations de l'incertitude épistémique (lié à un manque de connaissances ou à la présence de données subjectives) sur des problèmes pratiques d'analyse de risque. Deux types de problème sont étudiés:

1. l'estimation de **mesures d'importance** de composants en présence d'incertitudes épistémiques;
2. la **propagation d'incertitudes** dans un modèle employé pour analyser le risque inondation.

Le travail est focalisé sur l'**incertitude épistémique** qui affecte les paramètres d'entrée des modèles. Cette incertitude est représentée à l'aide de **distributions de probabilités** lorsque des données suffisantes sont disponibles pour permettre une analyse statistique, et par des **distributions de possibilités** lorsque l'information disponible provient d'experts, sous la forme d'estimations quantitatives imprécises ou de jugements.

Trois études de cas de complexité croissante sont présentées:

- ▷ un exemple pédagogique d'analyse de mesures d'importance concernant un système simple à trois composants, issu de la littérature;
- ▷ l'analyse de mesures d'importance pour un système auxiliaire d'alimentation en eau pour un réacteur nucléaire à eau pressurisée;
- ▷ une application en génie environnemental concernant la propagation d'incertitude dans un modèle hydraulique utilisé pour la conception d'une digue de protection contre les inondations.

About the authors

Enrico Zio est Professeur de Fiabilité et Analyse de Risque au *Politecnico di Milano*, et Directeur de la Chaire Systèmes complexes et défis énergétiques de l'École Centrale Paris & Supelec. Il est également chairman du *European Safety and Reliability Association* (ESRA).

Nicola Pedroni est maître de conférences au département Énergie du *Politecnico di Milano*. Sa recherche concerne les méthodes calculatoires avancées pour l'évaluation de la sécurité des systèmes industriels, en présence d'incertitude.

To cite this document

Zio et Pedroni (2013b). *Études de cas en propagation d'incertitude et mesures d'importance.* Numéro 2013-12 des *Cahiers de la Sécurité Industrielle*, Fondation pour une Culture de Sécurité Industrielle, Toulouse, France (ISSN 2100-3874). Disponible à l'adresse `http://www.FonCSI.org/fr/`.

Abstract

Title	Case studies in uncertainty propagation and importance measure assessment
Keywords	uncertainty, possibility theory, propagation, importance measures
Authors	Enrico Zio and Nicola Pedroni
Publication date	December 2013

The authors investigate the effects that different representations of epistemic uncertainty have on practical risk assessment problems. Two different application problems are considered:

1. the estimation of **component importance measures** in the presence of epistemic uncertainties;
2. the **propagation of uncertainties** through a risk flooding model.

The focus is on the **epistemic uncertainty** affecting the parameters of the models that describe the components' failures due to incomplete knowledge of their values. This epistemic uncertainty is represented using **probability distributions** when sufficient data is available for statistical analysis, and by **possibility distributions** when the information available to define the parameters' values comes from experts, in the form of imprecise quantitative statements or judgments.

Three case studies of increasing complexity are presented:

- ▷ a pedagogical example of importance measure assessment on a three-component system from the literature;
- ▷ assessment of importance measures for the auxiliary feed water system of a nuclear pressurized water reactor;
- ▷ an application in environmental modelling, with an analysis of uncertainty propagation in a hydraulic model for the risk-based design of a flood protection dike.

About the authors

Enrico Zio is Professor of Reliability, Safety and Risk Analysis at Politecnico di Milano and Director of the Chair in Complex Systems and the Energetic Challenge of École Centrale Paris & Supelec. He is also chairman of the *European Safety and Reliability Association* (ESRA).

Nicola Pedroni is an associate professor in the Energy Department of the Politecnico di Milano. His research concerns advanced computational methods for the reliability assessment of industrial systems in presence of uncertainties.

To cite this document

Zio and Pedroni (2013b). *Case studies in uncertainty propagation and importance measure assessment*. Number 2013-12 of the *Cahiers de la Sécurité Industrielle*, Foundation for an Industrial Safety Culture, Toulouse, France (ISSN 2100-3874). Available at http://www.FonCSI.org/en/.

Contents

Introduction

Context: uncertainty analysis in risk assessment

After having long been "swept under the carpet", uncertainty in risk analysis is now the subject of serious investigation, and is generally communicated to decision-makers and stakeholders as one of the outcomes of the risk analysis process. Two main families of uncertainty are considered: randomness due to inherent variability in the system behavior and imprecision due to lack of knowledge and information on the system. The former type of uncertainty is often referred to as objective, **aleatory**, stochastic whereas the latter is often referred to as subjective, **epistemic**, state-of-knowledge [Apostolakis 1990; Helton and Oberkampf 2004].

In current risk assessment practice, both types of uncertainties are represented by means of **probability distributions**. However, probabilistic representation of epistemic uncertainty is inappropriate when sufficient data is not available for statistical analysis, even if one adopts expert elicitation procedures to incorporate diffuse information into the corresponding probability distributions, within a subjective view of probability. Indeed, an expert may not have sufficiently refined knowledge or opinion to characterize the relevant epistemic uncertainty in terms of probability distributions [Helton and Oberkampf 2004].

As a result of the potential limitations of a probabilistic representation of epistemic uncertainty under limited information, a number of alternative representation frameworks have been proposed [Aven 2010, 2011; Aven and Steen 2010; Aven and Zio 2011; Flage et al. 2009], including fuzzy set theory [Klir and Yuan 1995], evidence theory [Ferson et al. 2003, 2004; Helton et al. 2007, 2008; Sentz and Ferson 2002], possibility theory [Baudrit and Dubois 2006; Baudrit et al. 2008; Dubois 2006; Dubois and Prade 1988b] and interval analysis [Ferson and Ginzburg 1996; Ferson and Hajagos 2004; Ferson and Tucker 2006; Ferson et al. 2007, 2010; Moore 1979]. A comparative analysis of these different representation frameworks has been published in the same collection as the present document [Zio and Pedroni 2013]. Possibility theory, in particular, may be the most attractive one for risk assessment, because of its representation power and its relative mathematical simplicity. It offers two measures of likelihood, namely possibility and necessity measures, that may be interpreted as lower and upper probabilities in the representation of imprecision in the experts' probability assignments.

Objectives of this document

In the present document, we assume a risk analysis setting in which the analyst uses probability distributions to describe aleatory uncertainty and a combination of probability and possibility distributions to describe epistemic uncertainty. The objective is to investigate the effects that different representations of epistemic uncertainty have on practical risk assessment problems involving decisions.

Two different application problems are considered:

1. The estimation of **component importance measures** in the presence of epistemic uncertainties [Baraldi et al. 2009b,a]. Importance measures are a technique used[1] to rank components or basic events in terms of their impact on system risk or reliability. During risk analysis of systems where many parameters are uncertain, they allow analysts to identify components whose impact on safety or reliability is significant, and focus effort on finding high-quality data for their reliability. They also allow analysts to identify "reliability bottlenecks", and thus focus maintenance spending on components that have a significant impact on system safety or reliability.
2. The **propagation of uncertainties** through a risk flooding model [Baraldi et al. 2012], in order to determine the dike level which is necessary to ensure a given flood return period, or the flood risk for a given dike level.

[1] The use of importance measures in risk analysis and maintenance optimization has been developed primarily in the nuclear power sector.

In this document, we focus on appropriate treatment of **epistemic uncertainty** affecting the parameters of the models that describe the components' failures or associated basic events (failure probabilities, failure rates, repair rates, unavailabilities), due to incomplete knowledge of their values **[Apostolakis 1990]**. Such epistemic uncertainties are represented here using:

- **probability distributions**, when sufficient data is available for statistical analysis **[Baraldi et al. 2009b]**;
- **possibility distributions**, when the information available to define the parameters' values comes from experts, in the form of imprecise quantitative statements or judgments **[Baraldi et al. 2009a]**.

Within this framework, the present work investigates

- how the epistemic (probabilistic or possibilistic) uncertainties in the model parameters can influence importance measures;
- how they can be accounted for in the ranking of the basic events or components;
- how the uncertainties in input probabilities can be propagated through a model to produce output uncertainties that can be communicated to decision-makers.

Three case studies of increasing complexity are presented.

Document structure

The document is organized as follows:

- Chapter 1 starts with a reminder of basic concepts about probability and possibility theories for epistemic uncertainty representation, then presents the methods used in the document to compare the importance of components or basic events in the presence of epistemic uncertainties. A technique developed by the authors for importance measure assessment in systems with large numbers of components is presented. It is based on the use of the Quicksort algorithm to avoid the combinatorial explosion of pairwise comparisons.
- Chapter 2 presents a simple case study of importance measure assessment with an pedagogical example involving a three-component system from the literature.
- Chapter 3 presents a case study of importance measure assessment for the auxiliary feed water system of a nuclear pressurized water reactor.
- Chapter 4 presents an application in environmental modelling, with a study of uncertainty propagation concerning the inputs of an hydraulic model for the risk-based design of a flood protection dike. The output variable of interest is the maximal water level reached by a river during the year.

Readers may be interested by three other documents by the same authors in the collection of the *Cahiers de la Sécurité Industrielle*:

- *Uncertainty characterization in risk analysis for decision-making practice* (CSI-2012-07), which provides an overview of sources of uncertainty which arise in each step of a probabilistic risk analysis;
- *Overview of risk-informed decision-making processes* (CSI-2012-10), which illustrates the way in which NASA and the US Nuclear Regulatory Commission implement risk-informed decision-making.
- *Literature review of methods for representing uncertainty* (CSI-2013-03), which provides an overview of probability theory, interval analysis and possibility theory and their use for risk analysis.

1

Methods used

1.1 Epistemic uncertainty representation

Two ways of representing the epistemic uncertainty affecting the parameters of risk models are described in the following: in particular, in § 1.1.1 the probabilistic framework is briefly recalled; in § 1.1.2 one of the non-probabilistic frameworks available in the literature, possibility theory, is introduced.

1.1.1 Probability theory

Let Ω be the space containing *all* the values that a given epistemically-uncertain parameter Y of interest can assume. In the discrete case, a *discrete* Probability Distribution Function (PDF) $d_Y(y)\colon \Omega \to [0,1]$ exists such that $\sum_{y\in\Omega} d_Y(y) = 1$; in the continuous case, a Probability Density Function (PDF) $p_Y(y)$ exists such that $\int_\Omega p_Y(y)\mathrm{d}y = 1$. For any measurable subset[1] A of Ω, the probability $P(A)$ of A is

$$\text{discrete case:} \quad P(A) = \sum_{y\in A} d_Y(y) \tag{1.1}$$

$$\text{continuous case:} \quad P(A) = \int_A p_Y(y)\mathrm{d}y \tag{1.2}$$

The probability $P(A)$ defined in equations (1.1) and (1.2) is required to have the following basic properties **[Helton and Oberkampf 2004]**:

1. if $A \in \Omega$, then $0 \le P(A) \le 1$;
2. $P(\Omega) = 1$;
3. if $A_1, A_2, \ldots, A_i, \ldots$, is a sequence of disjoint sets from Ω, then $P(\bigcup_i A_i) = \sum_i P(A_i)$;
4. $P(A) = 1 - P(\bar{A})$ (self-duality property): in words, the probability of a set A (*i.e.* $P(A)$) and the probability of its complement $\bar{A}$ (*i.e.* $P(\bar{A})$) must sum to one. Thus, specification of the likelihood of a set in probability theory also results in, or implies, a specification of the likelihood of its complement[2].

Finally, notice that in the continuous case the Cumulative Distribution Function (CDF) of Y is $F_Y\colon \Omega \to [0,1]$, defined from the PDF $p_Y(y)$ as follows:

$$F_Y(y) = P((-\infty, y]) = P(Y \le y) = \int_{-\infty}^{y} p_Y(t)\,\mathrm{d}t, \forall y \in \Omega \tag{1.3}$$

By way of example, let us assume that the epistemically-uncertain parameter Y is normal[3], *e.g.* $Y \sim N(5, 0.25)$: the corresponding PDF $p_Y(y)$ and CDF $F_Y(y)$ are shown in figure 1.1, left

[1] A *measure* on a set is a systematic way to assign a number to each suitable subset of that set, intuitively interpreted as its size.

[2] This property is *peculiar* to probability theory: in general, *less restrictive* conditions on the specification of likelihood are present in possibility theory (see § 1.1.2).

[3] Follows a normal (or gaussian) distribution.

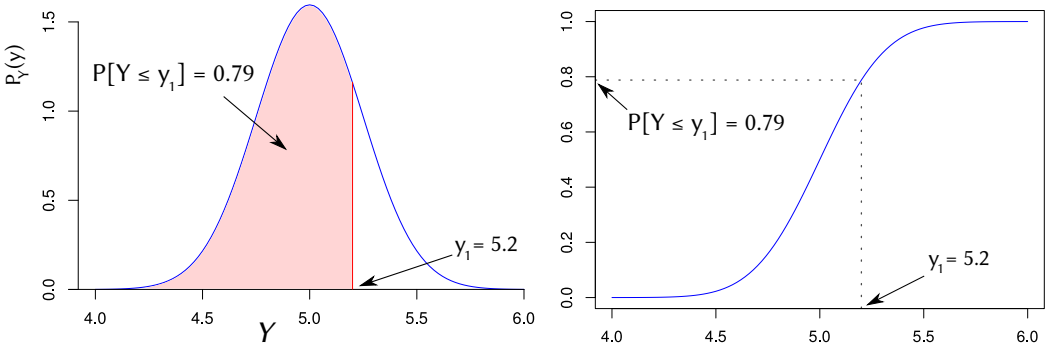


Figure 1.1 – *Probability density function, $p_Y(y)$ (left) and cumulative distribution function $F_Y(y)$ (right) of the normal random variable $Y \sim N(5, 0.25)$*

and right, respectively. The probability that the epistemically-uncertain parameter Y is lower than or equal to $y_1 = 5.2$, *i.e.* $P\{Y \leq y_1 = 5.2\} = \int_{-\infty}^{y_1=5.2} p_Y(y)\mathrm{d}y = 0.8$ is shown graphically in figure 1.1 (left) as the shaded area included between the PDF $p_Y(y)$ and the straight line $y_1 = 5.2$. Notice that this probability is equal to the value of the CDF $F_Y(y)$ in correspondence of $y_1 = 5.2$, *i.e.*, $F_Y(5.2) = 0.8$ (figure 1.1, right).

1.1.2 Possibility theory

The rationale for using possibility (instead of probability) distributions to describe epistemic uncertainty lies in the fact that a possibility distribution defines a *family* of probability distributions (bounded above and below by the so called possibility and necessity functions, respectively), thus it allows to account the expert's inability to specify a *single* probability distribution **[Baudrit and Dubois 2006; Baudrit et al. 2008; Dubois 2006; Dubois and Prade 1988a]**.

In possibility theory, uncertainty is represented by using a possibility function $\pi_Y(y)$. For each y in a set Ω, $\pi_Y(y)$ expresses the **degree of possibility** of y. When $\pi_Y(y) = 0$ for some y, it means that the outcome y is considered an impossible situation. When $\pi_Y(y) = 1$ for some y, it means that the outcome y is totally possible, *i.e.*, is just unsurprising, normal, usual **[Dubois 2006]**. This is a much weaker statement than when probability is 1.

The possibility function $\pi_Y(y)$ gives rise to probability bounds, upper and lower probabilities, referred to as necessity and possibility measures (N_Y, Π_Y). The possibility of a set A, $\Pi_Y(A)$, is defined by

$$\Pi_Y(A) = \sup_{y \in A} \{\pi_Y(y)\} \tag{1.4}$$

and the necessity measure $N_Y(A)$ is defined by

$$N_Y(A) = 1 - \Pi_Y(\overline{A}) = 1 - \sup_{y \notin A} \{\pi_Y(y)\} \tag{1.5}$$

Let $\mathcal{P}(\pi_Y)$ be a family of probability distributions such that for all sets A, $N_Y(A) \leq P(A) \leq \Pi_Y(A)$. Then,

$$N_Y(A) = \inf P(A) \tag{1.6}$$
$$\Pi_Y(A) = \sup P(A) \tag{1.7}$$

where *inf* and *sup* are with respect to all probability measures in $\mathcal{P}$. Hence the necessity measure is interpreted as a lower level for the probability and the possibility measure is interpreted as an upper limit. Referring to subjective probabilities, the bounds reflect the fact that the analyst is not able or willing to precisely assign her probability, and the bounds are the best she can do given the information available; in other words, she can only describe a subset of $\mathcal{P}$ which contains her probability **[Dubois and Prade 1988a]**.

A typical example of possibilistic representation is the following **[Anoop and Rao 2008; Baraldi and Zio 2008]**. We consider an epistemically-uncertain parameter y; based on its definition we know that the parameter can take values in the range $[4, 6]$ and the most likely value is 5: to represent this information a triangular possibility distribution on the interval $[4, 6]$ is used, with maximum value at 5, see figure 1.2.

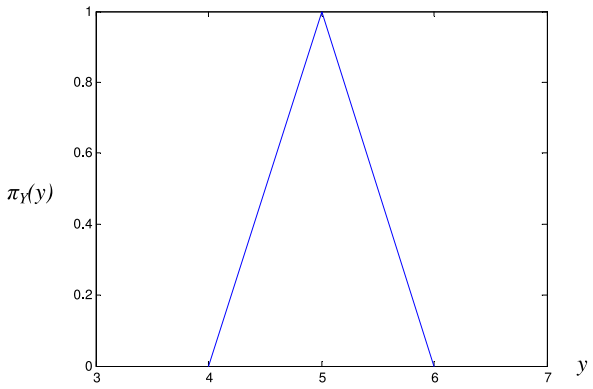


Figure 1.2 – *Possibility function for a parameter Y epistemically-uncertain on the interval [4,6], with maximum value at 5*

α-cut sets

DEFINITION

A fuzzy set is a collection of objects with various degrees of membership. It is sometimes useful to focus on those elements that have at least some minimal degree of membership α. For every $\alpha \in [0, 1]$, a given fuzzy set A yields a crisp set A_α which contains those elements of the universe which have membership of A at at least α level:

$A_\alpha = \{y: \pi_Y(y) \geq \alpha\}$, for $0 \leq \alpha \leq 1$.

Note that an α-cut of a fuzzy set is not a fuzzy set; it is a crisp set.

For example, $A_{0.5} = [4.5, 5.5]$ is the set of y values for which the possibility function is greater than or equal to 0.5. From the triangular possibility distribution in figure 1.2, we can conclude that if A expresses that the parameter lies in the interval $[4.5, 5.5]$, then $0.5 \leq P(A) \leq 1$.

From (1.5) we can deduce the associated cumulative necessity/possibility measures $N_Y(-\infty, y)$ and $\Pi_Y(-\infty, y)$ as shown in figure 1.3. These measures are interpreted as the lower and upper limiting cumulative probability distributions for the uncertain parameter y.

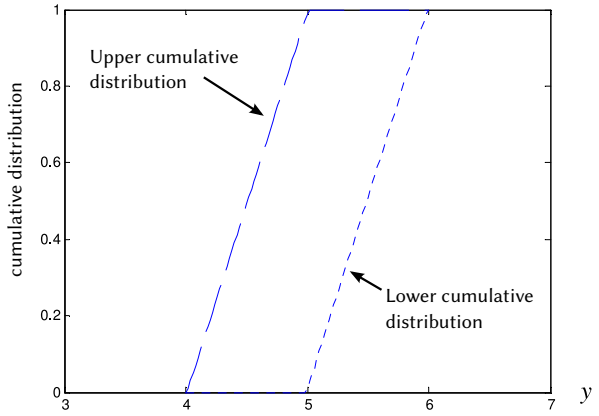


Figure 1.3 – *Bounds for the probability measures for the possibility function in figure 1.2*

1.2 Importance measures

Importance measures

DEFINITION

Importance Measures (IMs) are used to define the relative importance of the components of a system, according to some criteria. In a risk analysis context, they can be used to rank the contributions of components' failures or of basic events to the system risk **[Birnbaum 1969]**. An importance measure assigns a numerical value between 0 and 1 to each system component, where 1 indicates the highest importance.

Importance measures can be used for **risk-informed ranking** (*i.e.*, to arrange in order of risk importance) or categorizing (*i.e.*, to allocate into groups, according to pre-set risk criteria) components or, more generally, basic events in a plant risk model, to guide the plant design and Operation & Maintenance (O&M) **[Hoare 1962; Modarres 2006]**. With respect to maintenance, the objective is to focus the efforts on what is risk-important, while relaxing the activities on the low-significance groups of components, provided that at most only a small risk increase results, still confidently within the limits of acceptability **[Apostolakis 1990; Hoare 1962; Modarres 2006]**.

Different IM definitions can be used to address different aspects of reliability and risk (Fussel-Vesely-FV, Criticality, Risk Achievement Worth-RAW and Risk Reduction Worth-RRW) **[Apostolakis 1990; Youngblood 2001]**.

We address the case in which the models of the failure behaviors of the components rely on parameters which are poorly known, and thus tainted with imprecision.

Generally speaking, uncertainties may affect the parameters of the probability distributions that describe the components' stochastic failure behaviors. A number of techniques have been developed and widely investigated in the literature (*e.g.*, **[Helton and Oberkampf 2004; Baudrit et al. 2008; Kentel and Aral 2004, 2005, 2007; Möller 2004; Möller and Beer 2004, 2008; Möller et al. 2003, 2006]**) to address this case; they allow propagation of the uncertainties from the distributions' parameters to the probabilities $Q_j, j = 1...n$, of occurrence of the (basic) events of interest (for instance, from the component failure rate of an exponential distribution to its reliability at a given time instant). In other (simpler) cases, uncertainties may directly affect the probabilities $Q_j, j = 1...n$, of the basic events (component reliabilities or availabilities at a given time instant, failure probabilities of on-demand components).

These epistemic uncertainties will also affect the IMs of the components, and their consequent ranking. To handle the problem, three phases must be performed:

- **represent the imprecision** in the uncertain quantities of the failure models of the components;
- **propagate the uncertainty** onto the IMs values of the components $I_1, I_2, ..., I_j, ..., I_n$. Let $\underline{I} = (I_1, I_2, ..., I_j, ..., I_n)$ be the vector of importance measure values;
- **define a ranking** of the values $I_1, I_2, ..., I_j, ..., I_n$.

The first phase of representing uncertainty can be done in various ways, depending on the type, quality and quantity of the available information **[Zio and Pedroni 2013; Dubois et al. 1996; Helton and Oberkampf 2004]**. Generally speaking, when sufficiently informative data are available, uncertainties can be righteously represented and propagated within the probability theory framework (see § 1.1.1) **[Baraldi and Zio 2010; Baraldi et al. 2009b]**; in the opposite case, very often the only available information comes from experts, and is of ambiguous and qualitative nature: the uncertainty associated to this information is more naturally captured by other representation techniques such as possibility distributions (see § 1.1.2). That is, a distribution pair $H_{Q_j} = [N_{Q_j}, \Pi_{Q_j}]$ is introduced for every uncertain variable $Q_j, j = 1...n$, which represents either the unavailability or unreliability of the j-th component at a given time instant, for $j = 1...n$.

In the second phase, the epistemic uncertainties are propagated through the function $g(\underline{Q}) = g(Q_1, Q_2, ..., Q_j, ..., Q_n)$ that links the input variables $Q_j, j = 1...n$, to the output vector $\underline{I} = (I_1, I_2, ..., I_j, ..., I_n)$. In particular, $g(\underline{Q}) = g(\Phi(\underline{Q}))$, where $\Phi(\underline{Q})$ is the **structure function** of the system, which incorporates all the causal relations among the quantities Q_j that lead to the occurrence of the top event of interest (*i.e.*, system failure). When epistemic uncertainty is represented by probability distributions, the uncertainty propagation may rely, *e.g.*, on standard Monte Carlo Simulation (MCS) **[Kalos and Whitlock 1986]**; instead, when epistemic

uncertainty is represented by possibility distributions, the uncertainty propagation may rely on the relatively simple rules of fuzzy arithmetic **[Bojadziev and Bojadziev 1995]**, and provides as an output a couple of possibility and necessity measures Π_{I_j} and N_{I_j}, for $j = 1...n$.

The relevant contribution of the present work is related to the third issue: sorting the vector I. Practical procedures for ranking the components based on their uncertain IMs are proposed and discussed. This allows to concentrate the maintenance efforts on the components for which an improvement in maintenance results in a greater improvement of system performance (its reliability, risk, *etc.*). In § 1.3, a method is proposed to sort probabilistic uncertain quantities **[Baraldi et al. 2009b]**; § 1.4 describes an adaptation of the method to the case of possibilistic epistemic uncertainties **[Baraldi et al. 2009a]**.

1.3 Comparing the importance of components in presence of probabilistic epistemic uncertainties

In § 1.3.1, a probabilistic exceedance measure for the pairwise comparison of components' importance is proposed. In § 1.3.2, in order to extend the method to systems with large numbers of components, an empirical procedure for successive ranking is introduced to avoid the combinatorial explosion of pairwise comparisons **[Baraldi et al. 2009b]**.

1.3.1 A probabilistic exceedance measure for the pairwise comparison of components' importance

The aim of this section is to present a method for comparing the importance of two components A and B of an hypothetical system in presence of epistemic uncertainty on the components performance parameters (reliabilities, failure rates, repair rates *etc.*), which propagate through the system model leading to uncertainties in the system performance (*e.g.*, its reliability). In this scenario, importance measure calculations should reflect these uncertainties and so should the ranking.

With respect to uncertainty representation, in general when sufficiently informative data are available, probabilistic distributions are used. For simplicity of illustration, uniformly distributed uncertainty is assumed to be affecting directly the IMs of components A and B. Table 1.1 reports the ranges of the IMs distributions while figure 1.4 shows the corresponding distributions.

	Lower limit l	**Upper limit u**
A	0.0141	0.0155
B	0.0020	0.0178

Table 1.1 – *Parameters of the uniform distributions of uncertainty in the importance measures for components A and B*

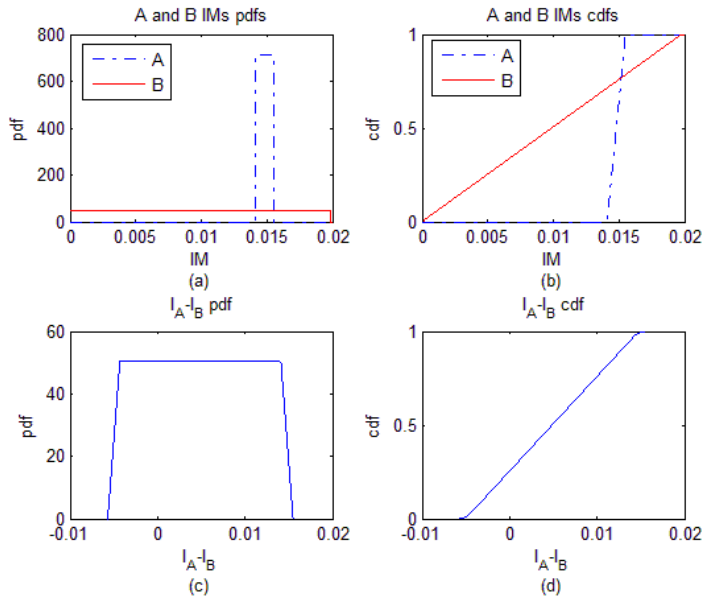


Figure 1.4 – *Probability density functions (pdfs) and cumulative distribution functions (cdfs) of the random variables I_A and I_B (a and b) and $I_A - I_B$ (c and d) in case of IMs with uniformly distributed uncertainties* [Baraldi et al. 2009b].

Looking at the distributions of the importance measures of A and B (denoted I_A and I_B, respectively) one may observe that the IM of component B (I_B), is significantly more uncertain than that of component A (I_A) but the expected value of I_A, $\mathbb{E}[I_A]$ is greater than that of B, $\mathbb{E}[I_B]$. On the other hand, there is a range in which the I_B quantiles are larger than the I_A ones. For example, if one were to perform the ranking based on the IMs 95^{th} quantile values, the conclusion would be that component B is more important than A, contrarily to what would be happen if the ranking were based on the expected values.

The drawback of comparing the expected values or specific quantiles lies in the loss of information about the distribution. With reference for example to figure 1.4, the fact that the 95^{th} quantile of I_A (0.015) is lower than that of I_B (0.017) only means that the point value which I_A is lower than with probability of 0.95 is lower than the analogous point value for I_B; the full information on the actual difference between the distributions of I_A and I_B does not play any role.

A natural way to give full account of the difference between the distributions of I_A and I_B is to consider the random variable (rv) $I_A - I_B$ whose pdf and cdf are shown in figure 1.4c and figure 1.4d, respectively. The details of their analytical expressions are given in Appendix A. In order to establish whether component A is more important than B, one can consider the probability $r_{AB} = 1 - F_{AB}(0)$ (where F is the cumulative distribution function) that I_A is greater than I_B; for example, in the present case $r_{AB} = P(I_A > I_B) = 1 - F_{AB}(0) = 0.81$, which means that with high probability component A is more important than B.

To decide on the relative importance of the two components A and B, it is necessary to fix a threshold $T \in [0.5, 1]$ on the r_{AB} value such that if r_{AB} is larger than T, then A is more important than B, otherwise no conclusion can be given. Obviously, the lower the threshold, the higher the risk associated with the decision.

Note however that the choice of a simple-valued threshold has some limitations when considering multiple components. For example, if the IMs of three components A, B and C are such that their differences all fall very close to T, it could happen that $I_A > I_B$, $I_B > I_C$ and $I_C > I_A$. Moreover, r_{AB} could fall very close to T, in which case no robust conclusion can be given on the components' importance given the inevitable approximations and uncertainties related to the estimation of the IMs distributions.

These limitations can partially be overcome by referring the comparison to a **threshold range** $[T_l, T_u]$ in such a way that for the two components A and B [Baraldi et al. 2009b]:

- ▷ If $r_{AB} > T_u$, then A is more important than B;

- ▷ If $r_{AB} < T_l$, then B is more important than A;
- ▷ If $T_l < r_{AB} < T_u$, then A is equally important to B. In this case, different kinds of additional constraints/targets can guide the ranking order (costs, times, impacts on public opinion, *etc.*).

For further insights, it is of interest to relate the importance measures results obtained by the **probabilistic exceedance measure** $r_{AB} = P(I_A > I_B)$ to the standard deviations of the IMs distributions, σ_{I_A} and σ_{I_B}. Figure 1.5 shows the variation of r_{AB} for increasing values of the standard deviation σ_{I_B}, keeping fixed the mean values of I_A and I_B and the ratio $k = \sigma_{I_A}/\sigma_{I_B}$ for different values of k. In the extreme case of no uncertainties on the knowledge of I_A and I_B ($\sigma_{I_A} = 0$ and $\sigma_{I_B} = 0$), component A is more important than B and thus $r_{AB} = 1$. Increasing the standard deviation σ_{I_B} (and thus also σ_{I_A}, keeping the ratio k constant), as expected $r_{AB} = 1$ holds as long as the pdfs of I_A and I_B do not overlap, *i.e.* I_A and I_B are uncertain quantities but it is not uncertain that $I_A > I_B$. The higher the ratio k, the lower the set of points for which $r_{AB} = 1$. Finally, as the overlapping between pdfs increases r_{AB} decreases **[Baraldi et al. 2009b]**.

From the above considerations, it can be argued that uncertainties can affect the components importance rank order and that reduction of uncertainties might be needed, in certain cases and when possible, to decrease the risk associated with the safety decision. To effectively drive the reduction of uncertainty, sensitivity analysis may be used, leading to the introduction of Uncertainty Importance Measures (UIMs) to identify the contribution of the epistemic uncertainty in the components' performance parameters to the importance measures' uncertainty **[Borgonovo 2006]**.

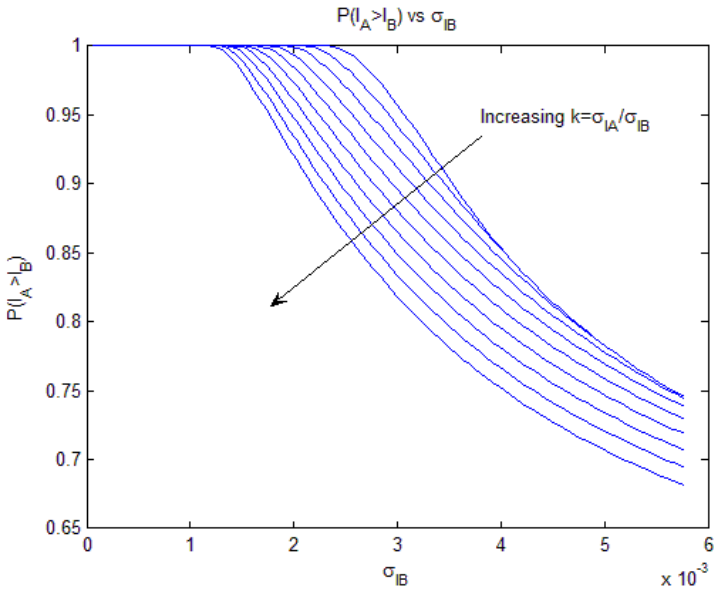


Figure 1.5 – r_{AB} *vs* σ_{I_B}, *keeping* $k = \sigma_{I_A}/\sigma_{I_B}$, $\mathbb{E}[I_A]$ *and* $\mathbb{E}[I_B]$ *constant* **[Baraldi et al. 2009b]**

1.3.2 An empirical procedure for component importance ranking in systems with large numbers of components

In the previous section, a probabilistic measure of exceedance between two random variables has been used to compare components' importance measures in presence of uncertainties. To extend the method to systems with a large numbers of components, a procedure for successive ranking must be introduced to avoid the combinatorial explosion of pairwise comparisons. The method proposed in this paper is an application of one of the most common sorting algorithms, Quicksort **[Hoare 1962]**. It is a divide-and-conquer algorithm, which relies on a partition of the elements based on a quantitative indicator of their 'size'. To partition the elements, it is required to choose one of them as pivot, *i.e.* as reference for moving all elements of size smaller before the pivot and all elements of size larger after it. In the resulting iterative partition procedure, the sublists of smaller and larger elements are recursively sorted.

In the case of interest here, the pivot element p is chosen as the component in the middle position of the components importance rank list obtained looking only at the mean values

of reliability/availability. The value of $r_{pj} = P(I_p > I_j)$ is then calculated for each component $j \neq p$ in the list and the pre-defined threshold range $[T_l, T_u]$ defines the rank order between p and j. The steps of the procedure are as follows **[Baraldi et al. 2009b]**:

1. Rank the components according to their IMs computed by considering the mean values of their reliability/availability parameters, *i.e.* without considering uncertainties.
2. Define the range $[T_l, T_u]$ of values of the probabilistic exceedance measure r_{pj}; for values r_{pj} in this range, it is not possible to decide whether $I_p > I_j$ or $I_p < I_j$ and this leads to consider components p and j as equally important, unless additional constraints/targets (costs, times, impacts on public opinion, *etc.*) allow the definition of an importance rank between the two.
3. Apply the Quicksort algorithm based on $r_{pj} = P(I_p > I_j)$:
   - 3.1 List the components in the rank order found in step 1;
   - 3.2 Choose the middle element of the list (sublist) as pivot element, p;
   - 3.3 For each j in the sublist compute the cdf F_{pf} of $I_p - I_j$ and evaluate $r_{pj} = 1 - F_{pj}(0)$:
     - ▷ If $r_{pj} > T_u$, then put j in the sublist of elements less important than p;
     - ▷ If $r_{pj} < T_l$, then put j in the sublist of elements more important than p;
     - ▷ If r_{pj} falls in $[T_l, T_u]$, then p is equally important to j.
   - 3.4 Append the sublist of less important elements to the right of p and the sublist of more important elements to the left of p;
   - 3.5 Recursively apply to each sublist steps 3.2-3.4 until no sublist with more than one element exists.

More details concerning the algorithm are given in Appendix B.

1.4 Comparing the importance of components in presence of possibilistic epistemic uncertainties

The ranking procedure discussed in the previous section is slightly modified to give due account to the fact that the IMs are not represented by probability distributions, but by possibility distributions, *i.e.*, families of probability distributions, whose upper and lower bounds are the possibility and necessity measures, respectively. Letting $\pi_{Q_s}(q_s)$ be the possibility distributions of the unreliabilities Q_s, $s = A, B$, of components A and B, the steps of the procedure are the following **[Baraldi et al. 2009a]**:

1. Compute the possibility distributions $\pi_{\Xi_{AB}}(\Xi_{AB})$ of the variable $\Xi_{AB} = \Xi_{AB}(I_A, I_B)$, defined as:

$$\Xi_{AB}(I_A, I_B) = \begin{cases} 1 & \text{if } \Delta_{AB} = I_A - I_B \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{1.8}$$

   where I_A and I_B are the IMs of components A and B, respectively, and $\Delta_{AB} = I_A - I_B$ is the difference between I_A and I_B.
   To do this:

   1.1. select a value of α on $[0, 1]$ and take as intervals of possible values of the unreliabilities of the components the cuts $[\underline{Q_s}, \overline{Q_s}]_\alpha = \left\{q_s | \pi_{Q_s}(q_s) \geq \alpha\right\}, s = A, B$;

   1.2. for every member I_s of the vector $\underline{I}$, calculate the smallest and largest values of $g(\Phi(\underline{Q}))$ (denoted by $\underline{g}_\alpha^{I_s}$ and $\overline{g}_\alpha^{I_s}$, respectively), when the elements of $\underline{Q}$ range within the intervals $[\underline{Q_s}, \overline{Q_s}]_\alpha$; that is, calculate:

$$\underline{g}_\alpha^{I_s} = \inf_{l, q_l \in [\underline{Q_l}, \overline{Q_l}]_\alpha} g(\underline{Q}) \tag{1.9}$$

$$\overline{g}_\alpha^{I_s} = \sup_{l, q_l \in [\underline{Q_l}, \overline{Q_l}]_\alpha} g(\underline{Q}) \tag{1.10}$$

   these are the lower and upper bounds, respectively, of the α-cut of the possibility distributions $\pi_{I_s}(i_s)$, $s = A, B$;

1.3. identify the bounds of the α-cut of the possibility distributions $\pi_{\Delta_{AB}}(\delta_{AB})$ of the variables $\Delta_{AB} = I_A - I_B$ (*i.e.*, the differences between the IMs of components A and B). These are given by $[\underline{g}_{\alpha}^{I_A} - \overline{g}_{\alpha}^{I_B}, \overline{g}_{\alpha}^{I_A} - \underline{g}_{\alpha}^{I_B}]$;

1.4. identify the α-cut of $\pi_{\Xi_{AB}}(\Xi_{\mathrm{AB}})$; in this regard, notice that $\pi_{\Xi_{AB}}(0) = \alpha$ if $\underline{g}_{\alpha}^{I_A} - \overline{g}_{\alpha}^{I_B} < 0$ and $\pi_{\Xi_{AB}}(1) = \alpha$ if $\overline{g}_{\alpha}^{I_A} - \underline{g}_{\alpha}^{I_B} \geq 0$.

1.5. repeat steps 1.1-1.4 for another value of α.

2. the relation order between the IMs of components A and B is established on the basis of the following rules:
   - ▷ if $\pi_{\Xi_{AB}}(0) \leq 0.3$, then component A is more important than B;
   - ▷ if $\pi_{\Xi_{AB}}(1) \leq 0.3$, then component B is more important than A;
   - ▷ components A and B are equally important in the other cases.

   This criterion is justified by the following considerations:
   - ▷ $\pi_{\Xi_{AB}}(\Xi_{AB}) = 1$ for at least one out of the two values of Ξ_{AB} (*i.e.*, 0 and 1), assuming that the distributions $\pi_{\Xi_{AB}}$ are normalized (*i.e.*, there must be at least one point of the Universe of Discourse (UoD) in which the distribution reaches 1);
   - ▷ The decision in favor of one out of the two components is taken only when the difference in their 'degrees of surprise' is large. A probabilistic interpretation of this, makes more clear the idea behind the criterion. For example, let us suppose that $\pi_{\Xi_{AB}}(0) = 0.2$ and $\pi_{\Xi_{AB}}(1) = 1$. Equation (1.5) allows to state that this is equivalent to $N_{\Xi_{AB}}(0) = 0$ and $N_{\Xi_{AB}}(1) = 0.8$, which can be interpreted as (equation (1.6)):

$$0.2 = \Pi_{\Xi_{AB}}(0) \geq P_{\Xi_{AB}}(0) = P(I_A \geq I_B) \geq N_{\Xi_{AB}}(0) = 0$$

$$1 = \Pi_{\Xi_{AB}}(1) \geq P_{\Xi_{AB}}(1) = P(I_A \leq I_B) \geq N_{\Xi_{AB}}(1) = 0.8$$

   To sum up, the probability that component B is more important than component A lies in the interval [0.8, 1], whereas the probability of the opposite case is a value between 0 and 0.2. In this situation, in which we are confident on the relevance of B with respect to A, it is reasonable to decide to concentrate the maintenance efforts on component B.

As in the probabilistic case (§ 1.3.2), when the number of components to be ranked is large (*i.e.*, » 2), the Quicksort algorithm **[Hoare 1962]** has to be used to avoid the combinatorial explosion of pairwise comparisons (see Appendix B for more details).

2

A three-component system

The first case study is an artificial example involving a three-component system of literature.

The system, sketched in figure 2.1, is made up of a series of two nodes: the first is constituted by two components in parallel logic, the second by a single component. Each of these components is characterized by a single valued reliability (column 2 of table 2.1). The values of some common IMs are reported in columns 3-6 of table 2.1. In the following, the discussion is limited to the Birnbaum IM but the reasoning remains exactly the same for the other IMs **[Baraldi et al. 2009b,a]**.

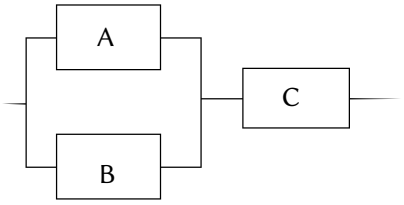


Figure 2.1 – *System Reliability Block Diagram*

	Reliability	Birnbaum	F-V	Criticality	RAW	RRW
A	0.985	0.009	0.002	0.001	1.094	1.001
B	0.990	0.014	0.002	0.001	1.141	1.001
C	0.905	0.9999	0.999	0.998	10.5	634

Table 2.1 – *Components' reliability and importance measures*

Let us now assume that the components' (un)reliabilities are epistemically-uncertain: probabilistic and possibilistic representations of such uncertainties are given in § 2.1 and § 2.2, respectively.

2.1 Probabilistic representation of epistemic uncertainty

Let us assume that the components are exponential, *i.e.* with constant failure rates λ_i, $i = A, B, C$ and that epistemic uncertainties affect their failure rates. The epistemic uncertainties in the failure rates are described by probability distributions described by the lognormal distributions of figure 2.2, left, with the parameters given in table 2.2 **[Baraldi et al. 2009b]**:

$$f_{\lambda_i}(\lambda_i) = \frac{e^{\frac{[ln(\lambda_i - \mu_i)]^2}{2\sigma_i^2}}}{\lambda_i \sigma_i \sqrt{2\pi}} \tag{2.1}$$

At each time instant t the reliability of component i is $r_i(t, \lambda_i) = e^{-\lambda_i t}$ with pdf (for $0 < \lambda_i < 1$) (figure 2.2, right):

$$f_{\lambda_i}(t, \lambda_i) = -\frac{e^{-\frac{[ln(-\frac{ln(\lambda_i)}{t}-\mu_i)]^2}{2\sigma_i^2}}}{\lambda_i ln(\lambda_i)\sigma_i\sqrt{2\pi}} \tag{2.2}$$

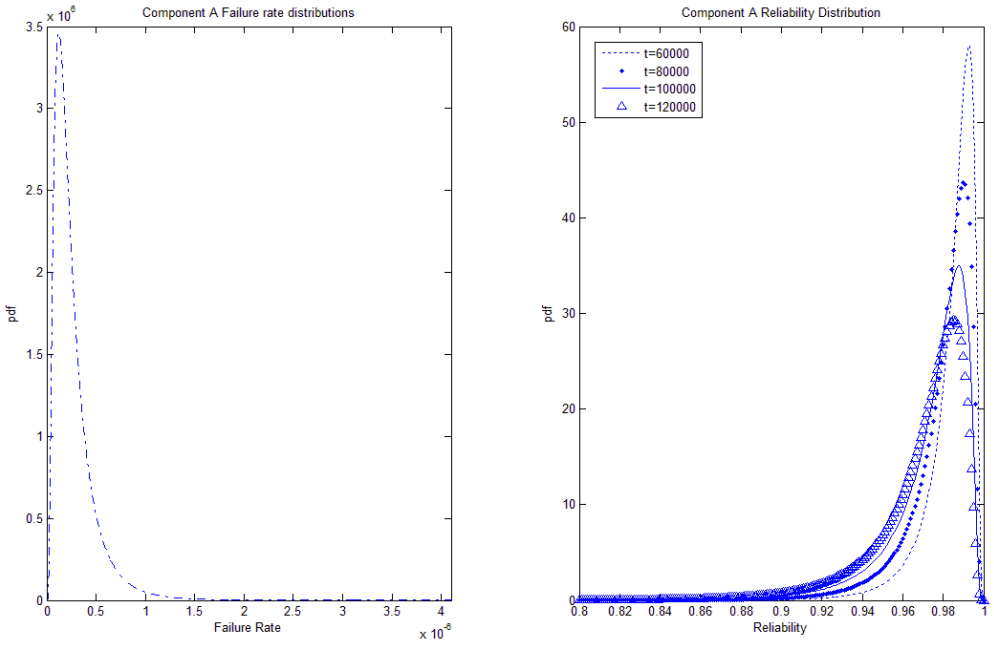


Figure 2.2 – *Lognormal distributions of the failure rate of component A (left) and corresponding pdfs of the reliability at different time instants (right)* [Baraldi et al. 2009b]

	Mean	Variance
A	1.00E-007	5.00E-08
B	1.50E-007	5.00E-08
C	1.00E-006	5.00E-07

Table 2.2 – *Parameters of the lognormal distributions of the components' failure rates*

The parameters of the distributions of the failure rates (table 2.2) have been chosen such that the mean values of the reliability at time $t = 10^5$ (in arbitrary units of time) are equal to the values in column 2 of table 2.1; in passing, notice that the standard deviations of the reliability at time $t = 10^5$ (in arbitrary units of time) are 0.005, 0.005 and 0.044 for components A, B and C, respectively. In figure 2.3 a and b, the pdfs of the failure rates and reliabilities at time $t = 10^5$ (in arbitrary units of time) are reported for all three components.

In spite of the simplicity of the considered system, finding the Birnbaum IM distributions by an analytical approach is impracticable. To overcome this difficulty, Monte Carlo sampling has been applied. The resulting distributions at the fixed time instant $t = 10^5$ are plotted in figure 2.3 c and d. It can be noted that the distribution of the IM of component C is displaced to larger values than that of components A and B, which leaves no doubt that the most important component is C, as expected from the structure of the system and the components' reliability values. As for the ranking of A and B, one must compute the r_{AB} measure (§ 1.3.1). The result obtained by Monte Carlo sampling is $r_{AB} = 0.23$, which with respect to $T_l = 0.3$ and $T_u = 0.7$ leads us to conclude that component B is more important than component A. Hence, the final component rank provided by the procedure proposed is CBA [Baraldi et al. 2009b].

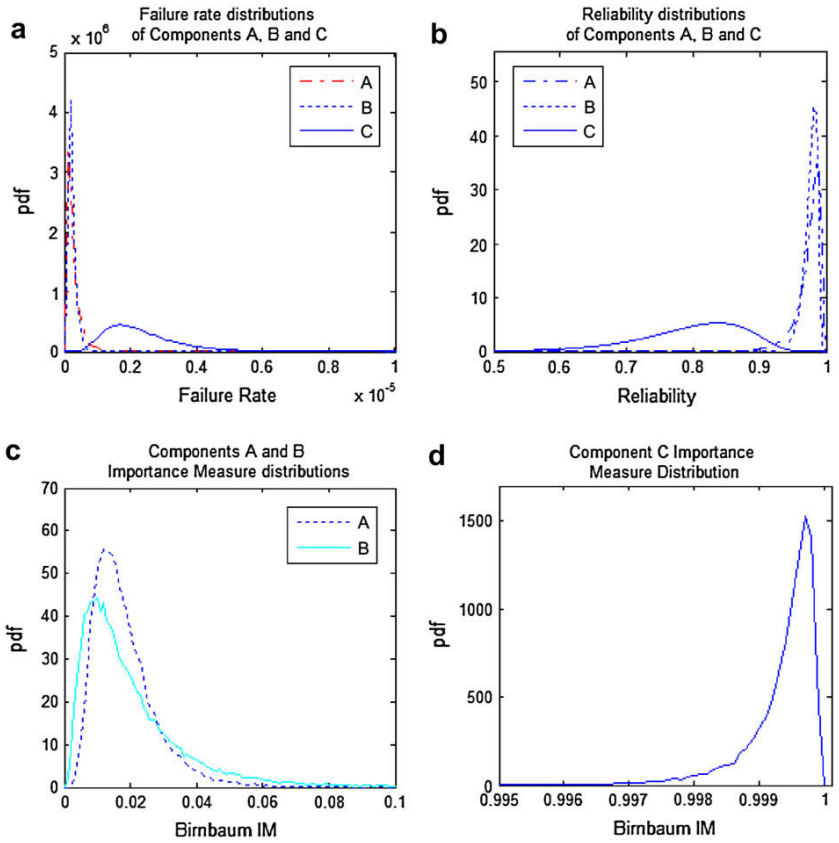


Figure 2.3 – *Pdfs of the failure rate (a) reliability (b) and Birnbaum IM (c and d) of the three components* [Baraldi et al. 2009b]

As a point of comparison, the procedure proposed in [Modarres 2006] has been applied (see Appendix C for details). The probability mass functions (pmfs) of the ranks of the three components obtained by Monte Carlo sampling of their uncertain failure rates are reported in figure 2.4. Notice that $r^*_{CA} = r^*_{CB} = 1 > r^*_{AB}$, which implies that component C is more important than both A and B also for this method. On the other hand, A is more important than B, given that the exceedance measure $r^*_{AB} = 0.52 \times (0.48{+}0.52) + 0.48{\times}0.52 = 0.77$. Notice, however, that if one considers $r^*_{BA} = 0.48 \times (0.52{+}0.48) + 0.52{\times}0.48 = 0.73$, B is more important than A; this shows that, in general, the exceedance measure $r^*_{ij} \neq 1 - r^*_{ji}$ is dependent on the choice of the pivot, and so is the final rank [Baraldi et al. 2009b].

It is interesting to investigate the relation between the two exceedance measures presented, r^*_{ij} and r_{ij}. Given that component C is the most important in all the M Monte Carlo samples ($P(R_C = 1) = 1$), the probability that A occupies a specific rank order between 2 and 3 is equivalent to the probability that B gains the only other rank order available, *i.e.* the probability mass value of rank order 2 of component A is the exceedance measure r_{AB} and, *vice versa*, the probability mass value of rank order 2 of component B is the exceedance measure r_{BA}.

Notice that the Birnbaum IM values in table 2.1, column 3, obtained neglecting uncertainties, would lead to the conclusion that B is more important than A.

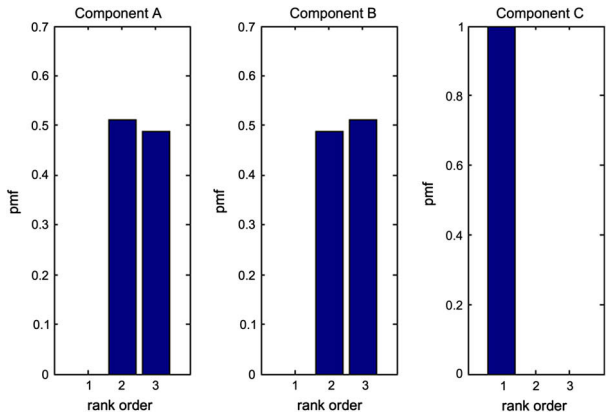


Figure 2.4 – *Probability mass functions of the rank orders of the three components* **[Baraldi et al. 2009b]**

2.2 Possibilistic representation of epistemic uncertainty

In the case study investigated in the previous section, the failure times of the components are assumed exponentially distributed with failure rates affected by epistemic uncertainties, which are modeled by lognormal distributions. Instead, in this section for the sake of simplicity, but with no loss of generality, we consider the case in which the epistemic uncertainties are described by means of possibility distributions $\pi_{Q_j}(q_j)$ and directly affect the values q_j of the components' unreliabilities $Q_j, j = A, B, C$ **[Baraldi et al. 2009a]**. To preserve the parallelism between the two frameworks (*i.e.*, probability and possibility theories), the possibility distributions $\pi_{Q_j}(q_j)$ are built coherently with the corresponding probability distributions $F_{Q_j}(q_j)$ described in the previous section. To do this, a number of techniques have been proposed in the literature (see **[Dubois 2006]** for a survey); one such technique, based on the Chebychev inequality

$$P\left(Q_j \in [q_j^* - a\sigma_j, q_j^* + a\sigma_j]\right) \geq 1 - \frac{1}{a^2} \quad \text{for } a \geq 1 \tag{2.3}$$

is adopted in this work **[Dubois 2006]**.

This inequality makes it possible to build a distribution-free possibilistic approximation of the unknown probability distribution of the value q_j, when all that is known is its mean value q_j^* and the standard deviation σ_j, and defines a bracketing approximation of symmetric intervals around the value q_j^*: notice the mean values q_j^* and standard deviations σ_j used in equation (2.3) are those of the probability distributions $F_{Q_j}(q_j)$ described in the previous section. The resulting possibility distributions $\pi_{Q_j}(q_j)$ are shown in figure 2.5 (a, b and c). The possibility and necessity measures Π_{Q_j} and N_{Q_j} associated with the possibility distributions $\pi_{Q_j}(q_j)$ of the unreliabilities in figure 2.5 (a, b and c) are shown in figure 2.5 (d, e and f), respectively, jointly with the Cumulative Distribution Functions (CDFs) $F_{Q_j}(q_j)$ associated to the pdfs in figure 2.3(b), for $j = A, B, C$. In this respect, notice that $F_{Q_j}(q_j) \in [N_{Q_j}, \Pi_{Q_j}]$, for every $j = A, B, C$. This proves that the CDFs are consistent with the corresponding possibility distributions; that is, $F_{Q_j}(q_j)$ belongs to the family $\mathbf{P}(\Pi_{Q_j}) = \left\{P_{Q_j}, \forall A \text{ measurable}, P_{Q_j}(A) \leq \Pi_{Q_j}(A)\right\} = \left\{P_{Q_j}, \forall A \text{ measurable}, N_{Q_j}(A) \leq P_{Q_j}(A)\right\}$ of all the probability distributions that are upper bounded by the possibility measure Π_{Q_j} and lower bounded by the necessity measure N_{Q_j}.

The uncertainties affecting the components' unreliabilities at a given time instant are propagated through the structure function of the three component system, and the possibility distributions describing the uncertainty on the importance of the components are found. Figure 2.6 shows the pairs $H_{I_j^B} = [N_{I_j^B}, \Pi_{I_j^B}]$ and the CDFs $F_{I_j^B}(i_j^B)$ of the Birnbaum IMs of the three components in figure 2.1 (notice that the CDFs $F_{I_j^B}(i_j^B)$ are associated to the probabilistic representation of epistemic uncertainty of § 2.1). Looking at this figure, it is intuitive to state that component C is more important of both components A and B, whatever the framework used to describe the uncertainties on the basic events. In fact, the possibility measure $\Pi_{I_C^B}$ (*i.e.*, the upper bound of the probability distributions) of the Birnbaum IM of component C, I_C^B, is far away from both the necessity measures $N_{I_A^B}$ and $N_{I_B^B}$ (*i.e.*, the lower bounds) of the

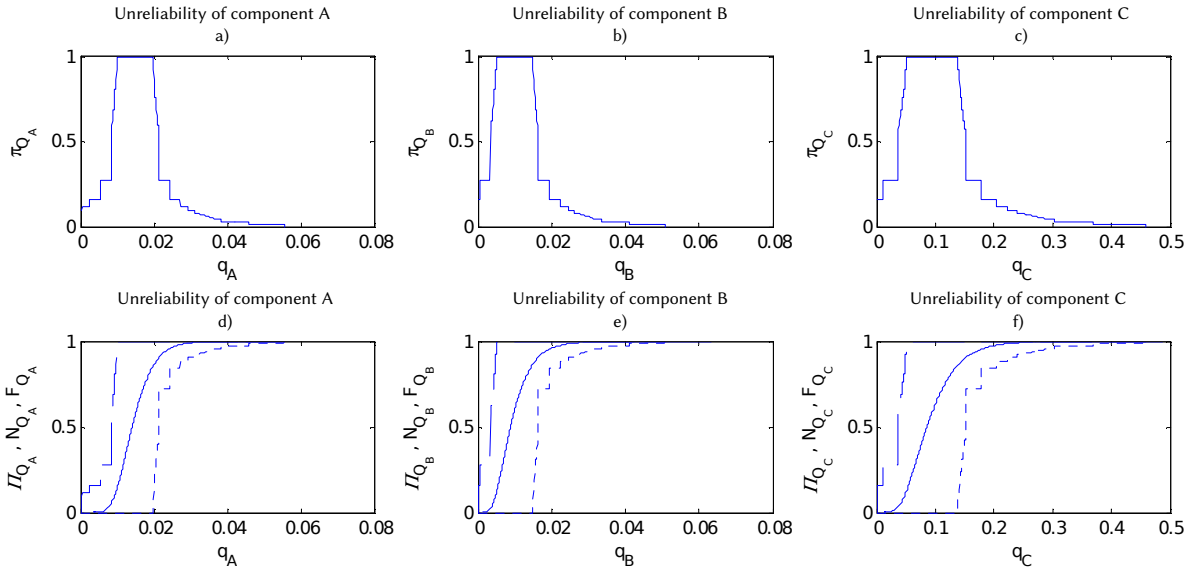


Figure 2.5 – *Possibility distributions of the unreliabilities of the components, a), b) and c) and their possibility measures, necessity measures and CDFs d), e) and f)* **[Baraldi et al. 2009a]**

IMs of components A and B, respectively: this brings the impossibility of A or B being more important than C **[Baraldi et al. 2009a]**.

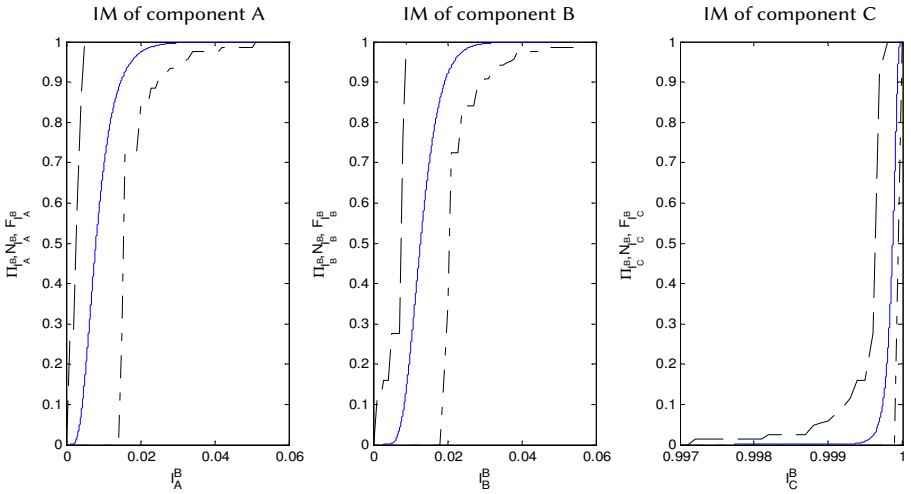


Figure 2.6 – *Possibility measure, necessity measure and CDF of the IMs of the components* **[Baraldi et al. 2009a]**

On the other hand, ranking the IMs of components A and B is not straightforward, as their possibility and necessity measures overlap each other: the procedure described in Section 3.2 is used to this aim. The results of the application of the proposed procedure are reported in figure 2.7 and figure 2.8: figure 2.7 shows the pairs $H_{\Delta_{jk}} = [N_{\Delta_{jk}}, \Pi_{\Delta_{jk}}]$ and the CDFs $F_{\Delta_{jk}}$ of the variables $\Delta_{jk} = I_j - I_k, j, k = A, B, C$ (*i.e.*, the differences between the IMs of the components); instead, figure 2.8 reports the possibility distributions $\pi_{\Xi_{kj}}(\xi_{kj})$ of the variables $\Xi_{kj} = \Xi_{kj}(I_k, I_j), j, k = A, B, C$, defined in § 1.4. Component C turns out to be the most important, whereas components A and B are equally important **[Baraldi et al. 2009a]**. Notice that in the probabilistic framework of the previous section, component B is classified as more important than component A. The fact that the results obtained in the probabilistic and possibilistic frameworks are different is unsurprising. In fact, as pointed out in **[Baraldi et al. 2009b,a]**, the difficulty in establishing a relation order between two uncertain quantities arises when their distributions overlap, that is, when there is a more or less extended zone of their UoD in which both probability densities are different from zero. In this case, the corresponding CDFs appear

close to each other or even intersect. Now, the CDFs of the unreliabilities in figure 2.5 d), e), f) are in the middle of the areas limited by the corresponding possibility and necessity measures; this entails that the closeness of the CDFs is less significant than that of their bounds. Thus, in some cases a distinction between the importance of two components may be possible in the probability framework, but not in the possibilistic one. Ultimately, this is due to the fact that the possibility distributions allow to represent and propagate a larger amount of uncertainty: within the probability framework one focuses just on one function of the infinitely many possible ones determined by a possibility distribution.

To sum up, the final ranking may be influenced by the choice of the framework in which the analysis is carried out, which basically depends on the available data.

The final ranking derived from the uncertain IMs depends also on how "open" the decision-maker wants to remain. Namely, in presence of a given amount of data, she may believe that resorting to the probability theory framework is justified. This leads to a final ranking which is expected to be capable of discriminating among the components based on their IMs. In the opposite case, the possibility distribution framework leaves the final ranking open to any distribution.

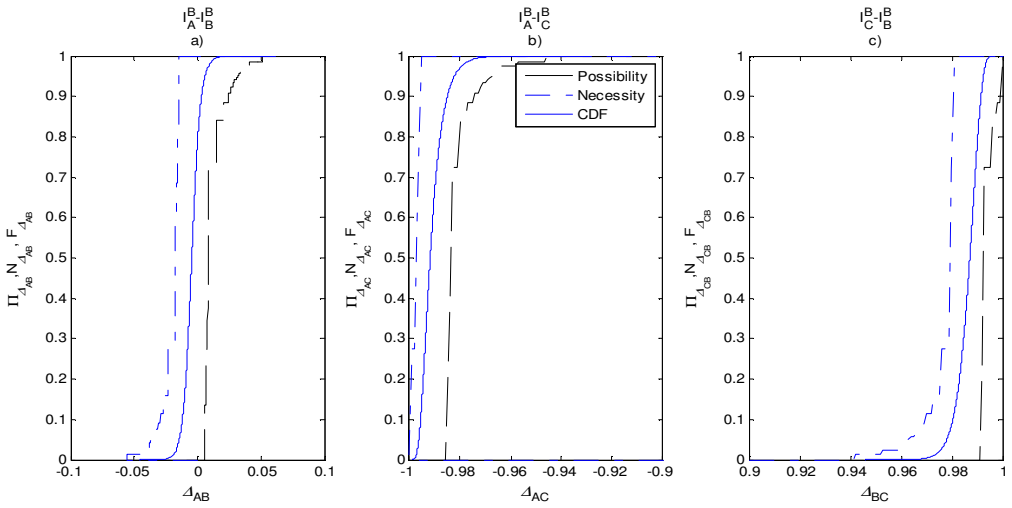


Figure 2.7 – *Comparison of the Importance measures: possibility measure, necessity measure and CDF of the variables* Δ_{kj}, $k, j = A, B, C$ *and* $k \neq j$

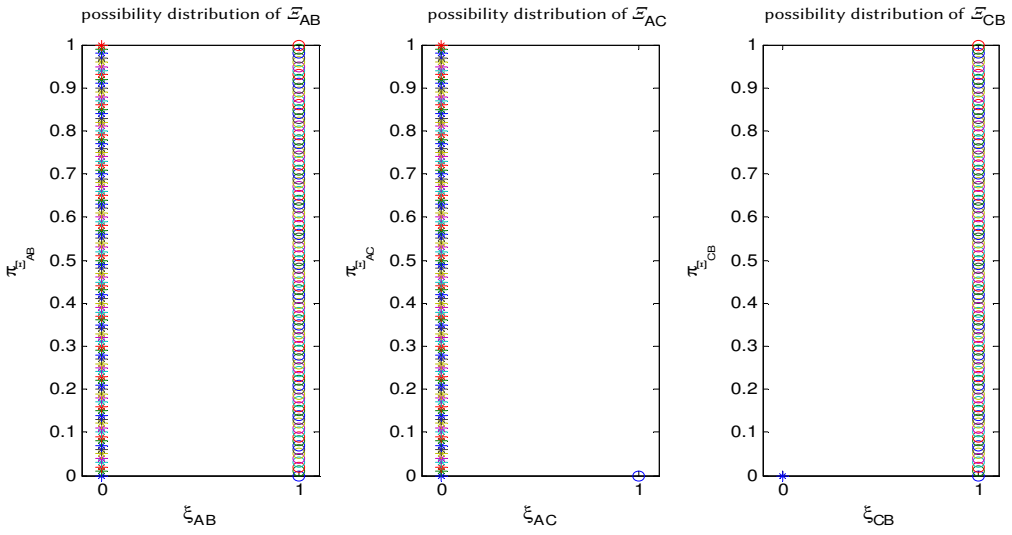


Figure 2.8 – *Comparison of the Importance measures: possibility distributions of the variables* Ξ_{kj}, $k, j = A, B, C$ *and* $k \neq j$

3

The auxiliary feedwater system of a nuclear pressurized water reactor

When the number of components in the system is large, the number of pairwise comparisons of their importance measures (IMs) needed for their importance ranking increases dramatically. This calls for a systematic procedure of analysis to efficiently perform the importance ranking. In this chapter, we will illustrate the calculation of an importance ranking on a larger system with more components.

Let us consider a simplified Auxiliary FeedWater System (AFWS) of a Pressurized Water Reactor (PWR) whose Reliability Block Diagram (RBD) is presented in figure 3.1. The case study is taken from **[Modarres 2006]**, where it is assumed that:

- ▷ all components are in standby mode;
- ▷ all components are periodically tested;
- ▷ the failure rates of all components are affected by epistemic uncertainty which is described by lognormal probability distributions.

With these assumptions, the average unavailability Q_j of each component $j = A, B, ..., N$, can be predicted by **[Modarres 2006; Zio 2007]**:

$$Q_j = \frac{1}{2}\lambda_j T_j^0 + f_j^r \times \frac{T_j^r}{T_j} + \frac{T_j^t}{T_j} \tag{3.1}$$

where λ_j is the failure rate (h^{-1}), T_j is the test interval (h), T_j^r is the average repair duration (h), T_j^t is the average test duration (h), f_j^r is the frequency of repair/test interval and $T_j^0 = T_j - T_j^r - T_j^t$ is the operating time (h), for any $j = A, B, ..., N$.

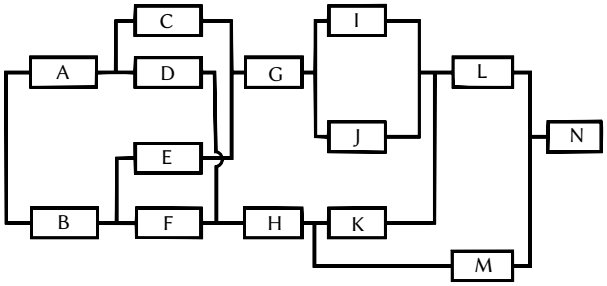


Figure 3.1 – *Reliability block diagram of the auxiliary feedwater system* **[Modarres 2006]**

Let us now assume that the components' failure rates (and, thus, the components' unavailabilities) are epistemically-uncertain: probabilistic and possibilistic representations of such uncertainties are given in § 3.1 and § 3.2.

3.1 Probabilistic representation of epistemic uncertainty

Table 3.1 contains the data relative to the lognormal distributions of the components' failure rates and to other operating characteristics **[Modarres 2006]**.

Name	Mean failure rate (μ_j)	Standard deviation of failure rates (σ_j)	Frequency of repair f_j^r	Average test duration $T_j^t(h)$	Average repair time $T_j^r(h)$	Test interval $T_j(h)$
A	1.00E-07	5.00E-08	9.20E-03	0	5	720
B	1.00E-07	5.00E-08	9.20E-03	0	5	720
C	1.00E-06	5.00E-07	2.50E-02	0	10	720
D	1.00E-06	5.00E-07	2.50E-02	0	10	720
E	1.00E-06	5.00E-07	2.50E-02	0	10	720
F	1.00E-06	5.00E-07	2.50E-02	0	10	720
G	1.00E-07	5.00E-08	7.70E-04	0	15	720
H	1.00E-07	5.00E-08	1.80E-04	0	24	720
I	1.00E-04	5.00E-05	6.80E-01	2	36	720
J	1.00E-04	5.00E-05	6.80E-01	2	36	720
K	1.00E-05	5.00E-06	5.50E-01	2	24	720
L	5.00E-07	2.50E-07	4.30E-03	0	10	720
M	3.00E-04	1.50E-04	1.50E-01	0	10	720
N	1.00E-07	5.00E-08	5.80E-04	0	5	720

Table 3.1 – *Failure data for the components of the auxiliary feedwater system*

In this case study, the analysis is illustrated with reference to the components' FV IMs, which are firstly computed with respect to their average unavailabilities (table 3.2, column 1). Notice that the fact of referring to average unavailabilities leads to no dependence of the probabilistic exceedance measure r_{ij}^* on time, unlike to the previous case study which made reference to the components' time-dependent reliabilities.

Let us now consider the epistemic uncertainties affecting the failure rates λ_i (table 3.1, column 3) and apply first the ranking procedure proposed in **[Modarres 2006]** (appendix C).

Figure 3.3 shows for each component of the system, the probability mass distributions of the rank orders obtained in step 3.4 of the procedure. Notice that these results are different from those reported in **[Modarres 2006]** with respect to both the mean rankings and the components' pmfs. These differences are due to the approximations used in **[Modarres 2006]** for computing the cut sets probabilities and the importance measures themselves. Table 3.2 reports the exceedance measures r_{ij}^* between the component i and the next two components in the ranking obtained by considering the average component unavailability (column 1). Following steps 4.1 and 4.2 of the procedure (with $[T_l, T_u] = [0.3, 0.7]$), the ranking reported in the last column of table 3.2 is obtained.

Then, the ranking procedure proposed in this work (*cf.* § 1.3) has been applied. Table 3.2 reports the probabilistic exceedance measures computed according to the proposed procedure between all possible pairs of components. However, notice that the application of the Quicksort algorithm requires the computation of only the measures reported in italic in table 3.3. Figure 3.7 shows the operations made by Quicksort: the bolded elements in the sublists are chosen as pivots and the arcs denote the result of the comparisons: components more important are put in the sublist on the left of the pivot, components less important on the right **[Baraldi et al. 2009b]**.

The first pivot is *H* (the element in the middle of the list) and the exceedance measures r_{Hj} for $j \neq H$ are computed; on this basis, the two sublists *MNLIJK* (more important components, left branch) and *GFCDEAB* (less important components, right branch) are created. Following, for example, the less important elements branch (right) of Figure 3.7, the middle element *D* is chosen as pivot of the sublist *GFCDEAB*; *F* has same rank order of *D*; *G*, *A* and *B* are placed together in the sublist of the more important elements and *C*, *E* in the lower importance sublist.

At the next iteration, A is chosen as pivot and the algorithm puts B in the same rank order as A in the lower importance sublist, whereas G goes into the sublist of higher importance. At the last iteration, C is taken as pivot and E is put in the sublist of the less important elements. The splitting and orders of this branch is completed because the base case of the recursion is met. The final ranking of this branch is (from left to right) G, (A, B), (D, F), C, E **[Baraldi et al. 2009b]**.

As for the overall final ranking of all the system elements, the proposed procedure (table 3.3, last column) and the procedure in **[Modarres 2006]** (table 3.2, last column) provide different results; this disagreement derives from the different numerical values of the exceedance measures (for example, $r_{MN} = 0.464$ while $r^*_{MN} = 0.776$). These differences are due to the fact that, for any i and j, r_{ij} depends only on the importance measures of i and j themselves whereas r^*_{ij} depends on the probability that a component occupies a specific order and thus also on the importance measures of the other components of the system **[Baraldi et al. 2009b]**.

It is also important to analyze the dependence of the final ranking from the choice of the pivot element in the ranking algorithm. The ranking procedure proposed in this work may encounter difficulties in cases characterized by, for example, three components i, j, k leading to $r_{ij}, r_{ik}, r_{jk} \approx T_u$ and $I_i = I_j, I_j = I_k$, as per figure 3.2. In this case, if component i is chosen as pivot the final rank is "i equally important to j" and (i, j) less important than k, whereas if j is chosen as pivot the three components are considered equally important and, finally, if k is chosen as pivot then i is less important than (j, k), which are equally important. However, this problem is partially overcome choosing the element in the middle of the list as pivot, in this case, i, j and k are considered equally important, and this seems the most reasonable choice. Notice that luckily these circumstances are expected to be very infrequent and can be handled case by case on the basis of other constraints/targets of interest **[Baraldi et al. 2009b]**.

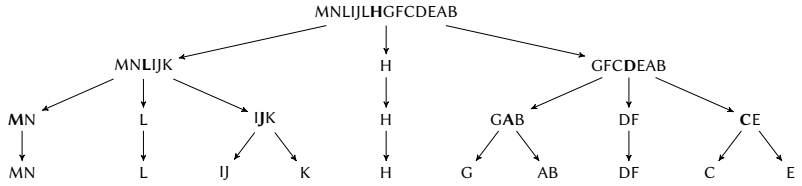


Figure 3.2 – *Quicksort procedure application example* **[Baraldi et al. 2009b]**

On the other hand, the final ranking is even more dependent from the choice of the pivot with the method proposed in **[Modarres 2006]**, which is characterized by an exceedance measure that does not satisfy $r^*_{ij} \neq 1 - r^*_{ji}$, as pointed out in § 2.1.

Finally, another parameter influencing the final results is the choice of the range $[T_l, T_u]$; for example, in the case study of the AFWS, considering [0.25, 0.75] instead of [0.3, 0.7] leads I, J, K to be equally important and so also H and G **[Baraldi et al. 2009b]**.

3.2 Possibilistic representation of epistemic uncertainty

Table 3.1 contains the values of the parameters of the lognormal distributions of the components' failure rates and other operating characteristics, which determine the values of means and standard deviations of the variables Q_j. Similarly to what was done in the simpler case study of chapter 2, the possibility distributions of the components' unavailabilities are built by putting these latter values into the Chebychev inequality (equation (2.3)). Figure 3.5 shows the corresponding possibility and necessity functions together with their CDFs.

Mean Rank	M	N	L	I	J	K	H	G	F	C	D	E	A	B	**Final rank**
M		0.776	1.000												M
N	0.603		0.886	0.917											N
L				0.805	0.804										L
I					0.683	0.994									*I,J*
J						0.976	1.000								*I,J*
K							1.000	1.000							*K*
H								0.892	0.989						*H*
G									0.969	0.973					*G*
F										0.651	0.618				*C,D,E,F*
C											0.583	0.632			*C,D,E,F*
D												0.662	0.999		*C,D,E,F*
E													1.000	1.000	*C,D,E,F*
A														0.751	*A*
B															*B*

Table 3.2 – *Components' FV importance measure rank orders obtained by considering the average unavailabilities (column 1); exceedance measures r^*_{ij} between the component i in the row and the following two i+1 and i+2 in the rank order of column 1 (columns 2-15); final rank obtained by the procedure proposed in* [Modarres 2006] *(last column)* [Baraldi et al. 2009b]

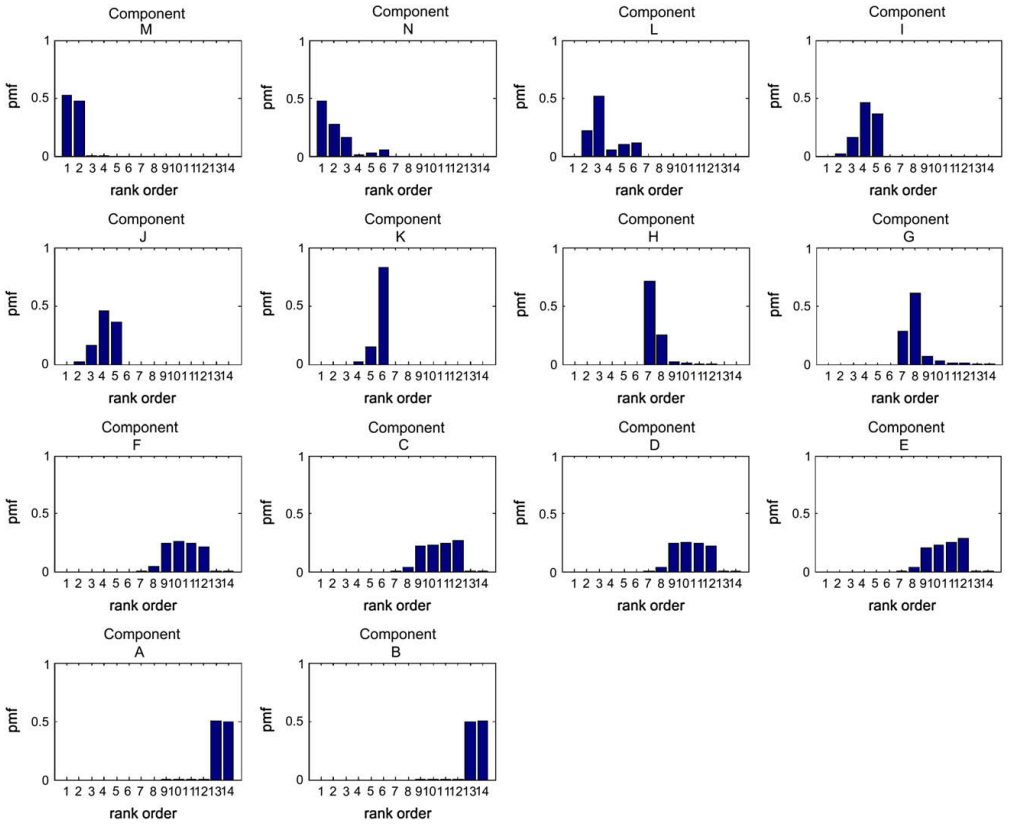


Figure 3.3 – *Probability mass functions of the components' rank orders* [Baraldi et al. 2009b]

Mean Rank	M	N	L	I	J	K	H	G	F	C	D	E	A	B	Final rank
M	-	*0.464*	*1.000*	1.000	1.000	1.000	*1.000*	1.000	1.000	1.000	1.000	1.000	1.000	1.000	*N,M*
N	-	-	*0.764*	0.949	0.949	0.949	*1.000*	1.000	1.000	1.000	1.000	1.000	1.000	1.000	*M,N*
L	-	-	-	*0.914*	*0.914*	*0.917*	*1.000*	1.000	1.000	1.000	1.000	1.000	1.000	1.000	*L*
I	-	-	-	-	*0.509*	*0.723*	*1.000*	1.000	1.000	1.000	1.000	1.000	1.000	1.000	*I,J*
J	-	-	-	-	-	*0.723*	*1.000*	1.000	1.000	1.000	1.000	1.000	1.000	1.000	*I,J*
K	-	-	-	-	-	-	*1.000*	1.000	1.000	1.000	1.000	1.000	1.000	1.000	*K*
H	-	-	-	-	-	-	-	*0.720*	*1.000*	*1.000*	*1.000*	*1.000*	*1.000*	*1.000*	*H*
G	-	-	-	-	-	-	-	-	1.000	*1.000*	*1.000*	1.000	*1.000*	*1.000*	*G*
F	-	-	-	-	-	-	-	-	-	*0.837*	*0.500*	0.923	0.004	0.004	*A,B*
C	-	-	-	-	-	-	-	-	-	-	*0.163*	*0.999*	*0.000*	*0.000*	*A,B*
D	-	-	-	-	-	-	-	-	-	-	-	*0.923*	*0.004*	*0.004*	*F,D*
E	-	-	-	-	-	-	-	-	-	-	-	-	0.000	0.000	*F,D*
A	-	-	-	-	-	-	-	-	-	-	-	-	-	*0.498*	*C*
B	-	-	-	-	-	-	-	-	-	-	-	-	-	-	E

Table 3.3 – *Components' FV IM rank orders obtained by considering the average unavailabilities (column 1); probabilistic exceedance measures r_{ij} between the component i in the row and the component j in the column (columns 2-15); components' FV IM rank orders obtained by applying the proposed probabilistic exceedance measure within the Quicksort algorithm (last column)* [Baraldi et al. 2009b]

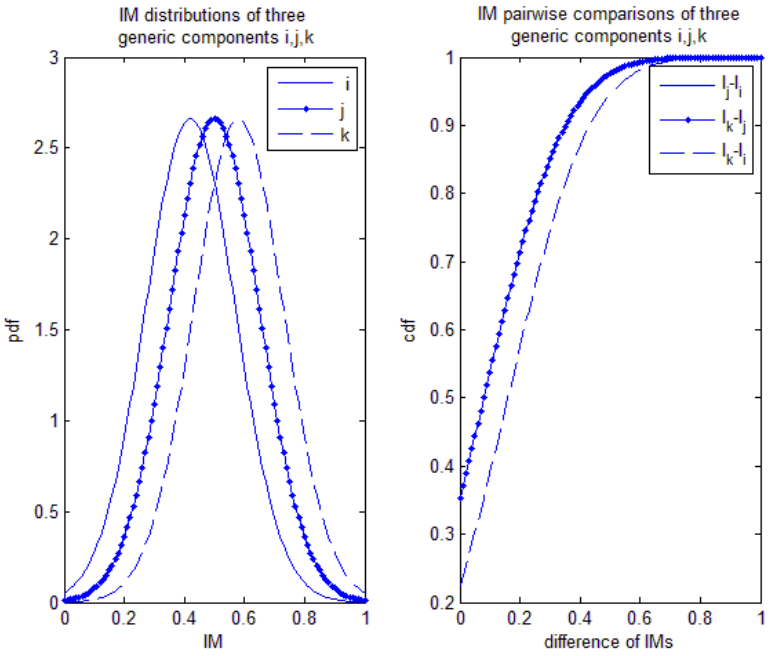


Figure 3.4 – *A particular case of importance measure comparison for three components i, j, k*

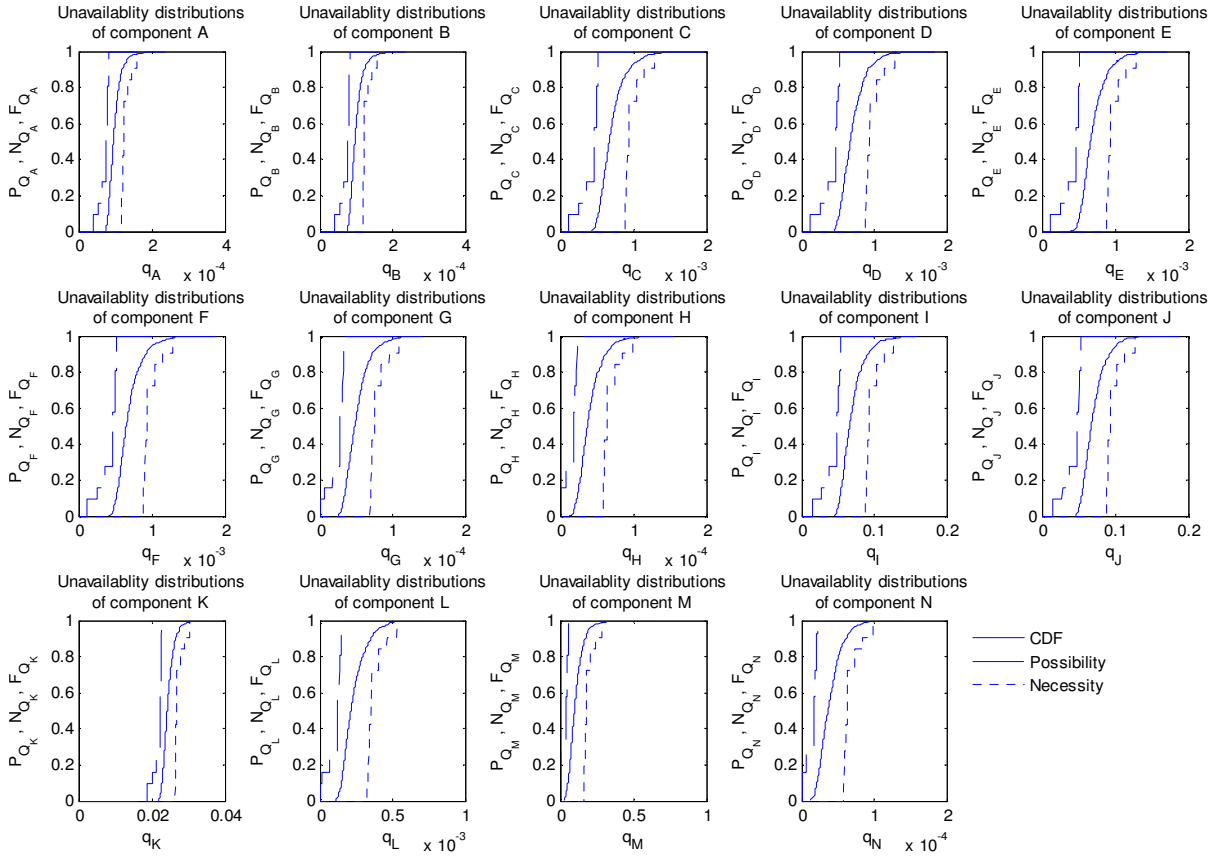


Figure 3.5 – *Possibility measures, necessity measures and CDFs of the unavailability of the components of the AFWS* [Baraldi et al. 2009a]

The uncertainties on components' unavailabilities $Q_A, ..., Q_N$ are propagated through the function $g(\Phi(\underline{Q}))$, and the Birnbaum IMs of the different components are obtained.

Figure 3.6 shows the CDFs, and the possibility and necessity measures of the Birnbaum IMs of the components of the AFWS. For visualization, these are reported in different scales. As expected, the largest measure is assigned to component *N*, which is a Single Point Failure (*i.e.*, its failure results in the loss of AWFS functionality). This state of affairs entails that the overall AWFS reliability is strongly sensitive to the improvement of the reliability of component *N*, and thus the maintenance actions on it are particularly effective.

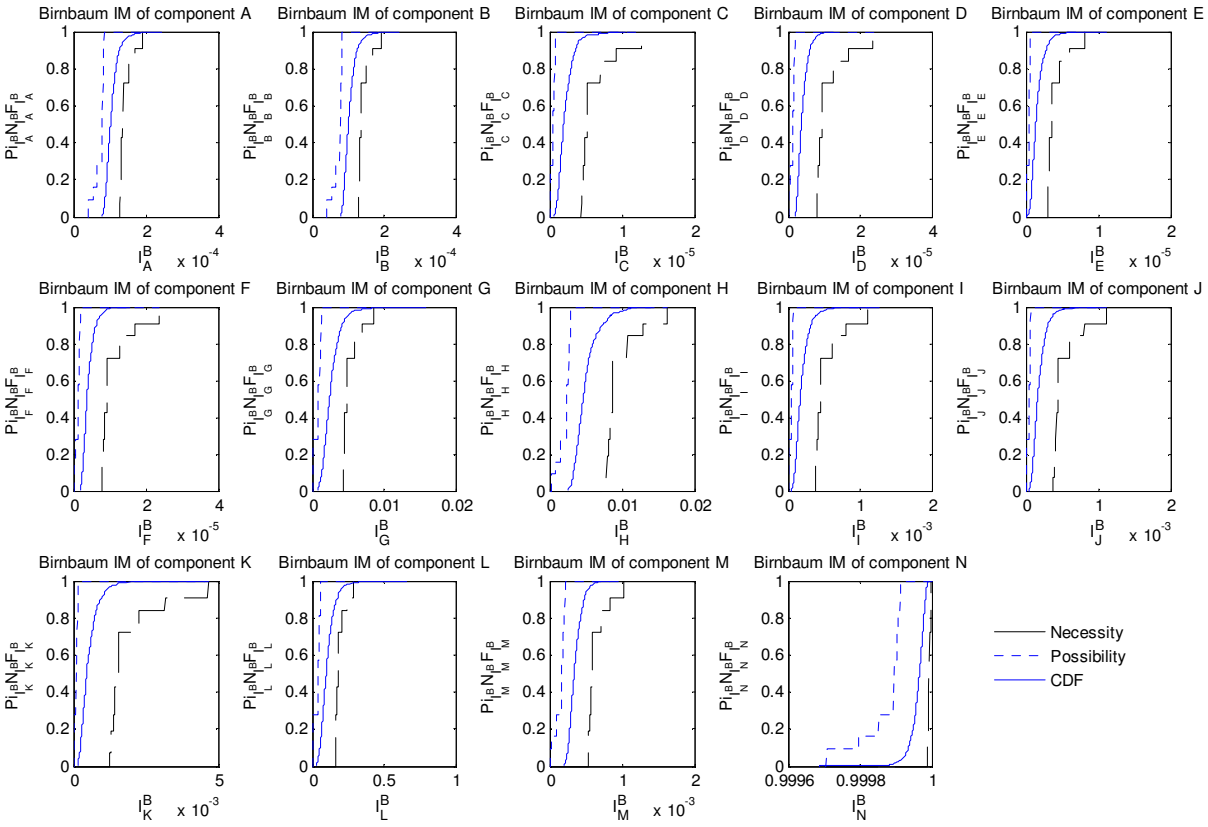


Figure 3.6 – *Possibility and necessity measures of the components' Birnbaum IMs* **[Baraldi et al. 2009a]**

The method proposed in § 1.4 is used to rank the components of the AFWS. Figure 3.7 shows the dynamics of the Quicksort algorithm in the considered case study, when the components are initially arranged in alphabetical order, and the pivot is always chosen as the central element of each sublist. At the first iteration, the pivot is *H* and two sublists are created: one containing the components that are equally or more important than the pivot (right branch, in this case it contains *L* and *N*) and the other with the less or equally important components (left branch, in this case it includes *ABCDEFGIJK*). Thus, *H* takes the third place instead of *L*, which currently occupies the second position, being more important than *H*. Notice that the pairwise comparisons also show that *H* is equally important than *G*, although the algorithm leaves *G* in its current position.

The sublist of more important components is then sorted: the comparison between *N* and *L* shows that the former is more important than the latter.

The less important elements branch contains *ABCDEFGMIJK*; its middle element, *F*, is chosen as pivot. The components *ABCDEFIJM* are more or equally important than *F*. In particular, this latter is equally important to *E*, *C* and *D*. With reference to the right sublist (more important components), *G* is the pivot element, and it swaps its position with *J*, that is, the importance of *J* is smaller than that of *G*. The algorithm proceeds as illustrated, and the final ranking is that reported in table 3.4. There is a first group of elements (*FCDE*) whose IMs are considerably smaller than those of the components of the second group (*ABIJM*). A further group (*GHK*) of components with similar importance has been identified; these are less important than *L* (the second most important component), which is considerably less important than *N* **[Baraldi et al. 2009a]**.

Ranking order	14-13-12-11	10-9-8-7-6	5-4-3	2	1
Components	*F D C E*	*A B I J M*	*G H K*	*L*	*N*

Table 3.4 – *Final ranking of the components' IMs (possibility theory framework)* **[Baraldi et al. 2009a]**

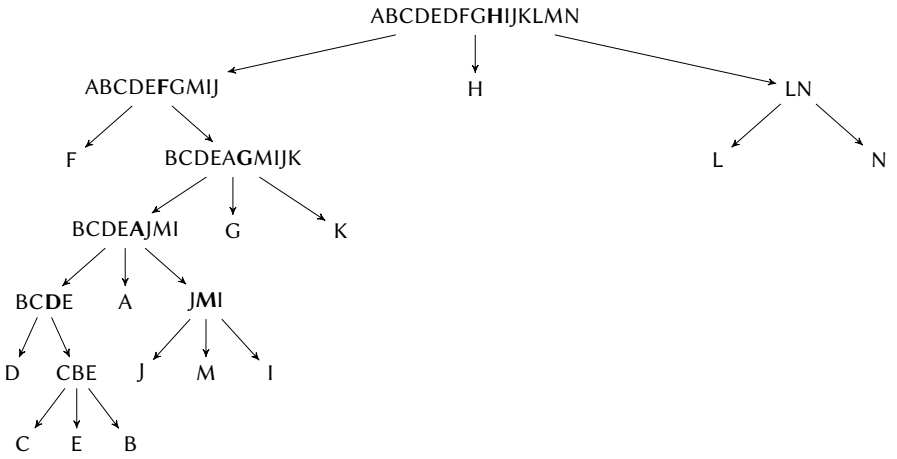


Figure 3.7 – *Steps of the Quicksort algorithm* [Baraldi et al. 2009a]

Notice that the final ranking depends on the initial arrangement of the components and on the choice of the pivot. For example, in the case investigated above, the Quicksort algorithm does not compare the importance of K and I; this comparison shows that they are equally important, so that the following relations hold $I_G^B = I_K^B$, $I_K^B = I_I^B$ and $I_I^B = I_j^B$. In this case, components G, K, I and $\mathcal{J}$ are considered equally important, even though the direct comparison between G and J shows that $I_G^B > I_j^B$. In turn, a more correct final ranking would be that reported in table 3.5.

An intuitive way to obtain a more "robust" final ranking is to run the sorting algorithm in correspondence of different choices of pivot selection policy and initial arrangement. Obviously, this may require large computation times. Notice also that, even if no attention is paid to the search for a more robust ranking, then the resulting maintenance decisions are neither "wrong" nor non-conservative; rather, the possibility of considering other decision criteria (cost, logistic aspects, *etc.*) is precluded.

Finally, the Quicksort algorithm is applied to the case in which the uncertainties on the basic events are described by probability distributions. The final ranking is reported in table 3.6. Some differences can be observed; that is, the components IMs are more distanced so that the group of components with similar IMs values are less numerous. Again, this is due to the fact that the overlaps between the CDFs are less significant than those of their bounds.

Ranking order	14-13-12-11	10-9-8-7-6-5-4-3	2	1
Components	$F\,D\,C\,E$	$A\,B\,I\,\mathcal{J}\,M\,G\,H\,K$	L	N

Table 3.5 – *Final ranking of the components' IMs (possibility theory framework), for a different choice of their initial arrangement* [Baraldi et al. 2009a]

Ranking order	14-13	12-11	10-9	8-7	6-5	5-4	3	2	1
Components	$C\,E$	$D\,F$	$A\,B$	$I\,\mathcal{J}$	$M\,K$	G	H	L	N

Table 3.6 – *Final ranking of the components' IMs (probability theory framework)* [Baraldi et al. 2009a]

4

Uncertainties in a flood risk model

This chapter illustrates uncertainty propagation on a more complex example, concerning the risk-based design of a flood protection dike.

4.1 Introduction

The growing concern about the environment and the potential risks associated with many human activities and new technologies have created increasing interest in Environmental Risk Assessment (ERA), which includes human health risk assessments, ecological or ecotoxicological assessments and specific industrial applications of risk assessment that examine end-points in people, biota or ecosystems [Fairman et al. 1998]. Environmental risk assessment is a critical, essential part of any decision-making process because it offers sound bases for assessing and ranking potential damages to the environment: as a consequence, the evaluation of environmental risk due to anthropic activities is an important step in mitigating their impact on natural resources and in recreating the co-evolutionary process between human and natural components of the environment [Darbra et al. 2008; Scandurra 1995].

Decision-makers of ecological policy and management require sound scientific information on the environmental risk associated with many different activities in order to arrive at and to justify their decisions: thus, there is a need to evaluate all potential risks that can cause environmental damage. This entails identifying all the potential actions, events and phenomena that may cause a damage to the environmental system of interest, quantifying the corresponding consequences and estimating their likelihood [Darbra et al. 2008; Fairman et al. 1998; Lein 1992].

Within this framework of analysis, the actions, events and physical phenomena that may cause damages to an environmental system, are in general described by complex mathematical models, which are then implemented in codes to simulate the behavior of the system of interest under various conditions [USNRC 2009; NASA 2010; EPA 2009].

In practice, *not all* the system characteristics can be fully captured in the mathematical model. This is due to the fact that i) many of the events and physical phenomena of interest are *random* in nature (the flooding of a river, an earthquake) and ii) the knowledge of the analyst about the phenomena involved is typically not complete. A complete environmental risk assessment usually requires a large amount of data. In some rare cases, extensive statistical data may be available and can contribute to an understanding of the frequency and the severity of the hazardous events of interest; however, it is very common that environmental data is scarce or qualitative, vague and imprecise [Darbra et al. 2008]. As a consequence, *uncertainty* is always present in the values of the *input parameters* and *variables* of the mathematical model: for example, in the models for groundwater risk assessment, elements of uncertainty lie in geological heterogeneity, physiological exposure parameters, and cancer potency [Maxwell et al. 1998]; in the models for the risk assessment of polluted sites, uncertainty may arise from the limited information that can be obtained from the contaminated sites due to technology limitations and costs of the analysis [Lehn and Temme 1996]; in the models for hydraulic risk assessment, uncertainty lies in hydraulic conductivity, specific yield, transmissivities, porosities, dispersivities and deoxygenation rate coefficients [Halkidis et al. 2009]. This input uncertainty propagates into variability in the *model output*.

In the work presented in this chapter, the uncertainties characterizing the inputs of an hydraulic model for the risk-based design of a flood protection dike have been analyzed and

propagated to the output variable of interest, the maximal water level reached by a river during the year. More specifically, the problem regards a dike that has to be built to protect a residential area close to the river. Two major issues have to be taken into account: i) the construction of a dike involves high building costs and annual maintenance costs; ii) the natural phenomenon of flooding is subject to uncertainties. Thus, the analyst has to evaluate different design options, accounting for the uncertainties **[Limbourg and de Rocquigny 2010]**.

For the treatment of uncertainty, we adopt the classical distinction between randomness due to inherent variability in the system behavior (aleatory, objective, stochastic uncertainty) and imprecision due to lack of knowledge and information on the system (epistemic, subjective, state of knowledge uncertainty) **[Apostolakis 1990; Helton and Oberkampf 2004]**.

With respect to the representation of the uncertainties, probability distributions are used to describe aleatory uncertainty **[Apostolakis and Kaplan 1981; Huanga et al. 2001; USNRC 2005, 2002, 2009; Maxwell et al. 1998]** and both probability and possibility distributions are used to describe epistemic uncertainty **[Baudrit et al. 2006, 2008; Dubois and Prade 1988a; Dubois 2006]**.

With respect to uncertainty propagation, a purely probabilistic approach is considered when epistemic uncertainty is described by probability distributions, whereas a 'hybrid' (*i.e.*, mixed probabilistic and possibilistic) approach is employed in the task of *jointly* propagating probabilistic and possibilistic uncertainties. The model for the risk-based design of a flood protection dike mentioned above **[Limbourg and de Rocquigny 2010]** is used as a benchmark for the comparison between the two approaches.

The remainder of this chapter is organized as follows. In section 4.2, the uncertainty representation frameworks (*i.e.*, probability and possibility theories) are briefly recalled, with particular emphasis on different methods used for building possibility distributions; in § 4.3, the details of the techniques applied for the joint propagation of aleatory and epistemic uncertainties are given; in § 4.4, the flood model considered for the uncertainty propagation task is presented and the results of the uncertainty propagation are reported and commented; in section 4.5, some conclusions are provided.

4.2 Uncertainty representation

In the work presented in this chapter, aleatory uncertainty is described by probability distributions and epistemic uncertainty is described by both probability and possibility distributions: the reader is referred to chapter 1 for basic details about probability theory (§ 1.1.1) and possibility theories (§ 1.1.2) for epistemic uncertainty representation.

In the following, the different approaches used for building possibility distributions are described in some detail.

4.2.1 Building possibility distributions

This section is mainly focused on approaches for constructing possibility distributions of the parameters/variables subject to epistemic uncertainty. In particular, in § 4.2.1, triangular possibility distributions are considered; in § 4.2.1, the use of Chebyshev inequality is illustrated; finally, in § 4.2.1, two methods for transforming a probability distribution into a possibility distribution are described based on the principle of maximum specificity and on the normalization of the probability density function.

Triangular function

Let us suppose that the analyst knows that an uncertain variable can take values in a given range $[a, b]$ and the most likely value is c. To represent this information a possibility distribution π_T can be taken as a triangle with basis determined by the range $[a, b]$ (*i.e.*, the absolute physical limits of the variable) and with vertex taken in correspondence of the most likely value c: in other words, the possibility distribution equals 0 in correspondence of the extreme values a and b of the physically allowable range and 1 in correspondence of the most likely value c. It has been shown that the family of probability distributions defined by a triangular possibility distribution π_T with range $[a, b]$ and vertex c contains all the probability distributions with support $I = [a, b]$ and mode c **[Baudrit and Dubois 2006]**.

Chebyshev inequality

If the analyst knows the mean μ and the standard deviation σ of the uncertain variable of interest, then the Chebyshev inequality **[Kendall and Stuart 1977]** can be used to construct a possibility distribution. Actually, the use of continuous possibility distributions for representing probability families heavily relies on probabilistic inequalities. Such inequalities provide probability bounds for intervals forming a continuous nested family around a typical value. This nestedness property leads to interpreting the corresponding family as being induced by a possibility measure. These bounds are usually used for proving convergence properties but, in this context, they can be used for representing knowledge. This is the case of the Chebyshev inequality, for instance.

The Chebyshev inequality

DEFINITION

The classical Chebyshev inequality **[Kendall and Stuart 1977]** defines a bracketing approximation on the confidence intervals around the known mean μ of a random variable Y, knowing its standard deviation σ. The Chebyshev inequality can be written as follows:

$$P(|Y - \mu| \leq k\sigma) \geq 1 - \frac{1}{k^2} \text{ for } k \geq 1 \tag{4.1}$$

The Chebyshev inequality defines a possibility distribution that dominates any density with given mean and variance: it allows to define a possibility distribution π by considering intervals $[\mu - k\sigma, \mu + k\sigma]$ as α-cuts of π and letting $\pi(\mu - k\sigma) = \pi(\mu + k\sigma) = \frac{1}{k^2}$. This possibility distribution defines a probability family $\mathcal{P}^{\mu,\sigma}(\pi)$ which has been proven to contain all distributions with mean μ and standard deviation σ, whether the unknown probability distribution function is symmetric or not, unimodal or not **[Baudrit and Dubois 2006]**.

Probability-possibility transformations

In this section, we consider transformations from probability distributions to possibility distributions. It is worth noting that in the transformation procedure (*i.e.*, going from probability to possibility) "some information is lost because there is a conversion from pointed-valued probabilities to interval-valued ones" **[Dubois et al. 1993]**.

Given the interpretation of possibility and necessity measures as upper and lower probabilities, a possibility distribution π induces a family $\mathcal{P}(\pi)$ of probability measures. There is not a one-to-one relation between possibility and probability, and transformations from a probability measure P into a possibility distribution π can only ensure that

1. $\mathcal{P}(\pi)$ includes P;
2. $\mathcal{P}(\pi)$ is selected according to some principle (rationale); *e.g.*, "minimize loss of information".

The following should be basic principles for such transformations **[Dubois et al. 1993]**:

- ▷ The **probability-possibility consistency principle**
  The family $\mathcal{P}(\pi)$ is formally defined as $\mathcal{P}(\pi) = \{P : \forall A \subseteq Y, P(A) \leq \Pi(A)\}$.
  It seems natural to require a transformation to select P from $\mathcal{P}(\pi)$ **[Dubois et al. 1993]**. This is referred to as the probability-possibility consistency principle, formulated as $P(A) \leq \Pi(A), \forall A \subseteq Y$.
- ▷ **Preference preservation**
  A possibility distribution π induces a preference ordering on Y, such that $\pi(y) > \pi(y')$ means that the outcome y is preferred to y'. A transformation should therefore satisfy $\pi(y) > \pi(y') \Leftrightarrow p(y) > p(y')$.

In the following, two methods for transforming a probability distribution into a possibility distribution are considered: the first one is based on the principle of maximum specificity; the second one on the normalization of the probability density function.

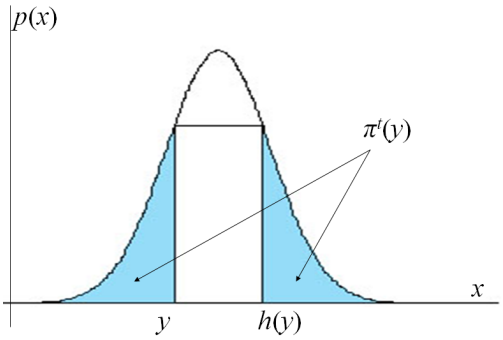


Figure 4.1 – *Transformation based on the principle of maximum specificity; the value of the possibility function π^t at y, $\pi^t(y)$, equals the shaded area*

The principle of maximum specificity The most specific possibility distribution π^t, or rather the minimum area under π, that dominates a given probability density p is given by:

$$\pi^t(y) = \pi^t(h(y)) = \int_{-\infty}^{y} p(x)\,\mathrm{d}x + \int_{h(y)}^{+\infty} p(x)\,\mathrm{d}x = F(y) + \overline{F}(h(y)) \tag{4.2}$$

where $\overline{F}(\cdot) = 1 - F(\cdot)$ and $h(y) = \max\{x: p(x) \geq p(y)\}$. For the sake of clarity, the transformation in (4.2) is graphically illustrated in figure 4.1.

It is interesting to observe that for this transformation:

$$N([y, h(y)]) = P\left([y, h(y)]\right) \leq \Pi([y, h(y)]) = 1$$

i.e., the transformation prescribes equality between the necessity of a given α-cut and the probability of the same α-cut.

The transformation applies to unimodal, continuous and support bounded probability densities p. Moreover this criterion is not necessarily suitable for the transformation of a subjective probability distribution reflecting an expert opinion.

Normalization of probability density The possibility distribution resulting from the transformation is given by the normalization of probability density, *i.e.*, $\mu_p = \frac{p(y)}{\sup\, p(y)}$. Note that the distribution resulting from this normalization (when taken to be a possibility distribution) does not in general adhere to the probability-possibility consistency principle **[Dubois and Prade 1980]**.

4.3 Propagation of aleatory and epistemic uncertainties through a model

Let us consider a model whose output is a function $Z = f(Y_1, Y_2, ..., Y_i, ..., Y_n)$ of n uncertain variables Y_i, $i = 1...n$, ordered in such a way that the first k are affected by purely aleatory uncertainty and the last $n - k$ by purely epistemic uncertainty. The aleatory uncertainty is described by probability distributions $p^{Y_j}(y_j), j = 1...k$; on the contrary, the epistemic uncertainty may be represented either by probability distributions, $p^{Y_l}(y_l), l = k + 1...n$, or by possibility distributions, $\pi^{Y_l}(y_l), l = k + 1...n$. A graphical representation of this framework for uncertainty modeling is given in figure 4.2.

If both aleatory and epistemic uncertainties are represented by probability distributions, a purely probabilistic approach **[Kalos and Whitlock 1986; Marseguerra and Zio 2002]** based on Monte Carlo simulation is embraced (§ 1.3); if the aleatory and epistemic uncertainties are represented by probability and possibility distributions, respectively, a Monte Carlo simulation and fuzzy interval analysis approach **[Baudrit et al. 2006; Baraldi and Zio 2008; Flage et al. 2010a,b]** is considered (§ 1.4).

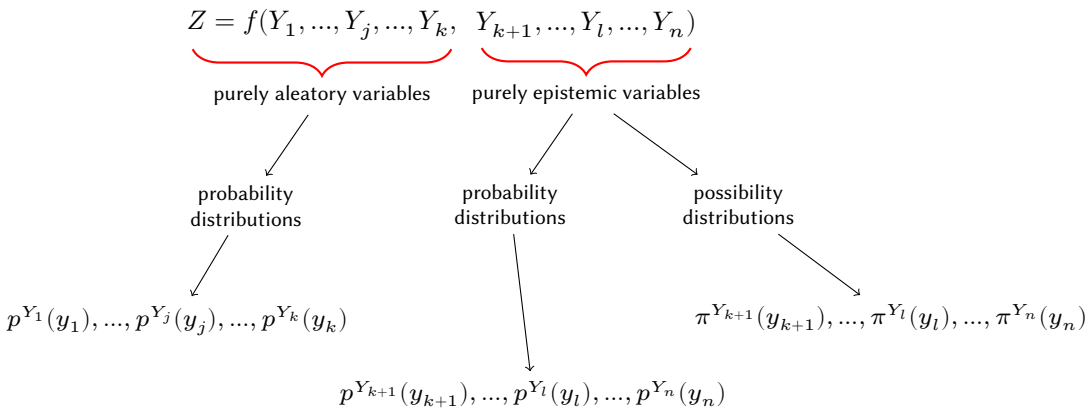


Figure 4.2 – *Scheme of the uncertainty modeling framework*

4.3.1 Purely probabilistic approach

In the purely probabilistic approach, all the input variables are considered probabilistic and their uncertainty is represented by probability distributions. The purely probabilistic approach is based on the Monte Carlo sampling of possible values of all the input variables from the corresponding probability distributions and the subsequent computation of the model output which correspond to the input values sampled **[Kalos and Whitlock 1986]**. This procedure is repeated a large number of times to collect different values of the model output in correspondence of different values of the input variables. These random realizations can be used to calculate quantities of interest, *e.g.*, the empirical cumulative distribution function of the model output.

4.3.2 Monte Carlo simulation and fuzzy interval analysis approach

In the Monte Carlo simulation and fuzzy interval analysis approach (hereafter also called hybrid approach), the epistemic uncertainty is represented in possibilistic terms. Therefore, the joint propagation of the aleatory and epistemic uncertainty can be performed by combining the Monte Carlo technique **[Kalos and Whitlock 1986]** with the extension principle of fuzzy set theory **[Zadeh 1965]** by means of the following two main steps **[Baudrit et al. 2006]**:

1. repeated Monte Carlo sampling of the random variables to process aleatory uncertainty;
2. fuzzy interval analysis to process epistemic uncertainty.

For the generic i^{th} k-tuple of values sampled by Monte Carlo, $i = 1...m$, an estimate of $Z = f(Y_1, Y_2, ..., Y_n)$ is obtained in terms of a fuzzy subset π_i^f. As a result of the m repeated samplings of the random variables, $Z = f(Y_1, Y_2, ..., Y_n)$ turns out to be represented as a fuzzy random variable (or random possibility distribution) in the sense of **[Gil 2001]**.

The operative steps of the procedure are (see figure 4.3 to figure 4.5):

1. set $i = 0$;
2. set $i = i + 1$;
3. sample the i^{th} realization $(y_1^i, ..., y_k^i)$ of the random variable vector $(Y_1, ..., Y_k)$;
4. set $\alpha = 0$;
5. select the α-cuts $A_\alpha^{k+1}, A_\alpha^{k+2}, ..., A_\alpha^n$ of the possibility distributions $(\pi^{Y_{k+1}}, ..., \pi^{Y_n})$ as intervals of possible values of the possibilistic variables $(Y_{k+1}...Y_n)$; by way of example, the α-cut A_0^{k+1} of level $\alpha = 0$ and the α-cut $A_{0.7}^{k+1}$ of level $\alpha = 0.7$ are shown in figure 4.3 and figure 4.4, respectively, with reference to the possibility distribution $\pi^{Y_{k+1}}$ of the uncertain variable Y_{k+1};

6. calculate the smallest and largest values of $f(y_1^i, ..., y_k^i, Y_{k+1}, ..., Y_n)$, denoted by $\underline{f}_\alpha^i$ and $\overline{f}_\alpha^i$ respectively, considering the fixed values $(y_1^i, ..., y_k^i)$ sampled for the random variables $(Y_1 ... Y_k)$ and all values of the possibilistic variables $(Y_{k+1} ... Y_n)$ in the α-cuts $A_\alpha^{k+1}, A_\alpha^{k+2} ... A_\alpha^n$ of their possibility distributions $(\pi^{Y_{k+1}} ... \pi^{Y_n})$, *i.e.*, $\underline{f}_\alpha^i = \min_{Y_{k+1} \in A_\alpha^{k+1} ... Y_n \in A_\alpha^n} f(y_1^i, ..., y_k^i, Y_{k+1}, ..., Y_n)$ and $\overline{f}_\alpha^i = \max_{Y_{k+1} \in A_\alpha^{k+1} ... Y_n \in A_\alpha^n} f(y_1^i, ..., y_k^i, Y_{k+1}, ..., Y_n)$;

7. take the extreme values $\underline{f}_\alpha^i$ and $\overline{f}_\alpha^i$ found in step 6. as the lower and upper limit of the α-cut of $f(y_1^i, ..., y_k^i, Y_{k+1}, ..., Y_n)$; by way of example, the α-cut A_0^f of level α = 0 and the α-cut $A_{0.7}^f$ of level α = 0.7 are shown in figure 4.3 and figure 4.4, respectively, with reference to the possibility distribution π_i^f of $Z = f(Y)$;

8. if α ≠ 1 then set α = α + Δα (*e.g.*, Δα = 0.05) and return to step 5. above; otherwise, obtain the fuzzy random realization (fuzzy interval) π_i^f of $Z = f(Y)$ as the collection of the values $\underline{f}_\alpha^i$ and $\overline{f}_\alpha^i$ for each α-cut; an example of fuzzy random realization (fuzzy interval) $\pi_i^f(z)$ of $Z = f(Y)$ is pictorially shown in figure 4.5;

9. if $i \neq m$ then go back to step 2. above, else stop the algorithm.

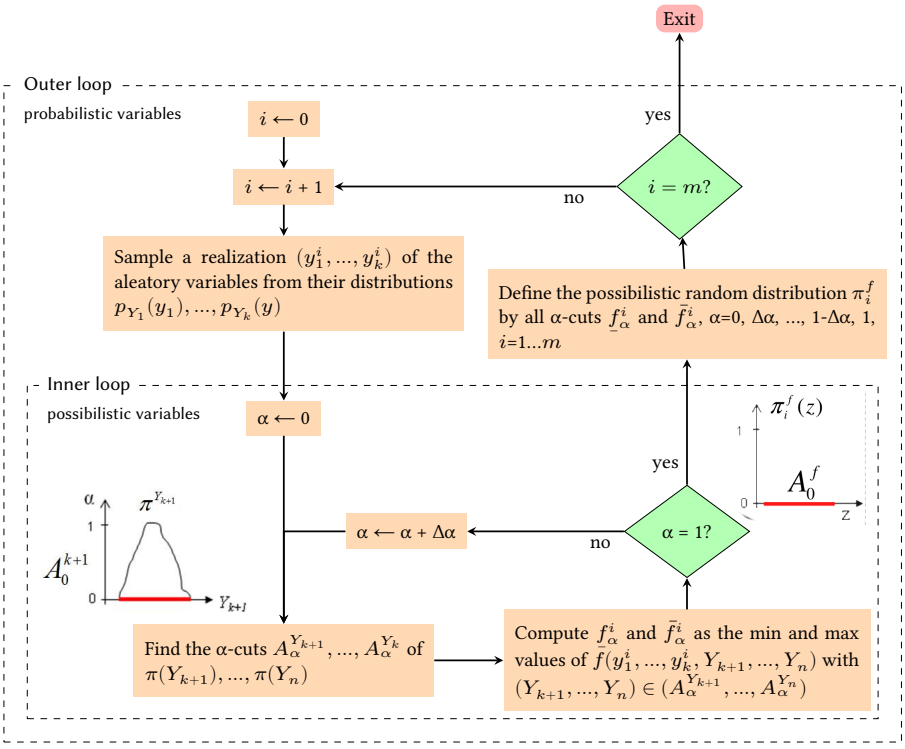


Figure 4.3 – *Scheme of the operative steps of the hybrid method: in evidence the selection of the α-cut A_0^{k+1} of level α = 0 of the possibility distribution $\pi^{Y^{k+1}}$ of variable Y^{k+1} and the computation of the α-cut A_0^f of level α = 0 of the possibility distribution π_i^f of $Z = f(Y)$*

The procedure is repeated for $i = 1...m$: at the end of the procedure an ensemble of realizations of fuzzy intervals is obtained for the output function $Z = f(Y)$, *i.e.*, $(\pi_1^f ... \pi_m^f)$ (figure 4.6 on the left).

For each set A contained in the universe of discourse U_Z of the output variable Z, it is possible to obtain the possibility measure $\Pi_i^f(A)$ and the necessity measure $N_i^f(A)$ from the corresponding possibility distribution $\pi_i^f(z)$, by:

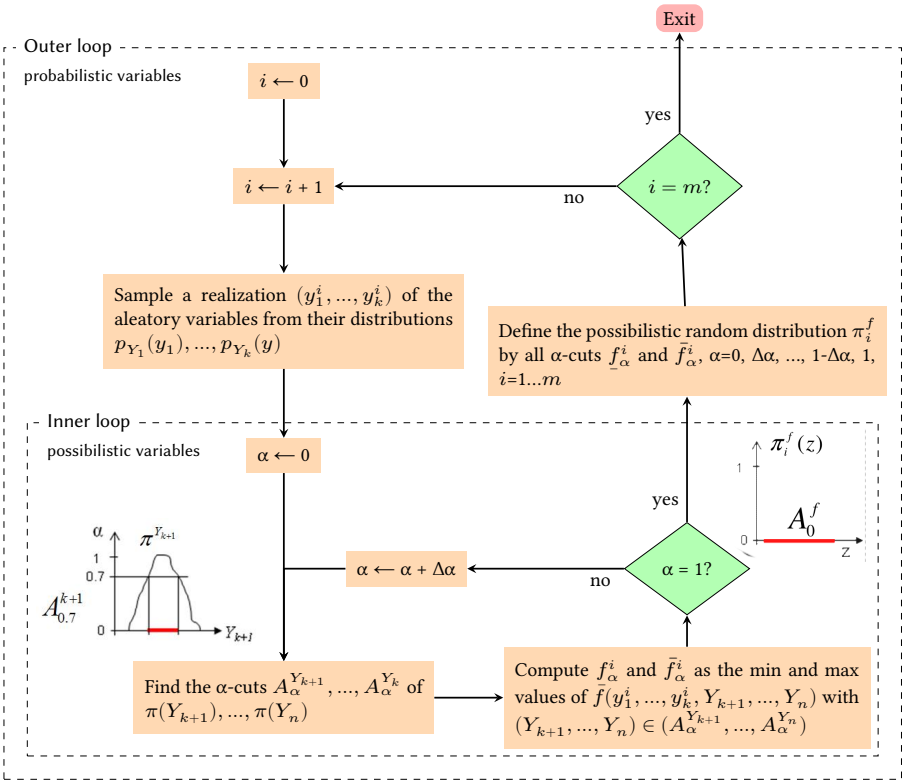


Figure 4.4 – *Scheme of the operative steps of the hybrid method: in evidence the selection of the α-cut $A_{0.7}^{k+1}$ of level $\alpha = 0.7$ of the possibility distribution $\pi^{Y^{k+1}}$ of variable Y^{k+1} and the computation of the α-cut $A_{0.7}^f$ of level $\alpha = 0.7$ of the possibility distribution π_i^f of $Z = f(Y)$*

$$\Pi_i^f(A) = \max_{z \in A} \left\{ \pi_i^f(z) \right\} \tag{4.3}$$

$$N_i^f(A) = \inf_{z \notin A} \left\{ 1 - \pi_i^f(z) \right\} = 1 - \Pi_i^f(\overline{A}) \quad \forall A \subseteq U_Z \tag{4.4}$$

The m different realizations of possibility and necessity can then be combined to obtain the belief $\mathrm{Bel}(A)$ and the plausibility $\mathrm{Pl}(A)$ for any set A, respectively **[Baudrit et al. 2006]**:

$$\mathrm{Bel}(A) = \sum_{i=1}^{m} p_i N_i^f(A) \tag{4.5}$$

$$\mathrm{Pl}(A) = \sum_{i=1}^{m} p_i \Pi_i^f(A) \tag{4.6}$$

where p_i is the probability of sampling the i^{th} realization $(y_1^i, ..., y_k^i)$ of the random variable vector $(Y_1 ... Y_k)$; in the present case of m realizations $p_i = 1/m$.

In this view, the likelihood of the value $f(Y)$ passing a given threshold z can then be computed by considering the belief and the plausibility of the set $A = (-\infty, z]$; in this respect, $\mathrm{Bel}\left(f(Y) \in (-\infty, z]\right)$ and $\mathrm{Pl}\left(f(Y) \in (-\infty, z]\right)$ can be interpreted as bounding, average cumulative distributions $\underline{F}(z) = \mathrm{Bel}\left(f(Y) \in (-\infty, z]\right), \overline{F}(z) = \mathrm{Pl}\left(f(Y) \in (-\infty, z]\right)$ **[Baudrit et al. 2006]**.

Let the *core* and the *support* of a possibilistic distribution $\pi^f(z)$ be the crisp sets of all points of U_Z such that $\pi^f(z)$ is equal to 1 and nonzero, respectively. Considering a generic value z of $f(Y)$, it is $\mathrm{Pl}\left(f(Y) \in (-\infty, z]\right) = 1$ if and only if $\Pi_i^f\left(f(Y) \in (-\infty, z]\right) = 1, \forall i = 1...m$, that is, for $z > z^* = \max_i \left\{ \inf(\mathrm{core}(\pi_i^f)) \right\}$. Similarly, $\mathrm{Pl}\left(f(Y) \in (-\infty, z]\right) = 0$ if and only if $\Pi_i^f\left(f(Y) \in (-\infty, z]\right) = 0 \ \forall i = 1...m$, that is, for $z \leq z_* = \min_i \left\{ \inf(\mathrm{support}(\pi_i^f)) \right\}$.

Finally, one way to estimate the total uncertainty on $f(Y)$ is to provide a confidence interval at a given level of confidence, taking the lower and upper bounds

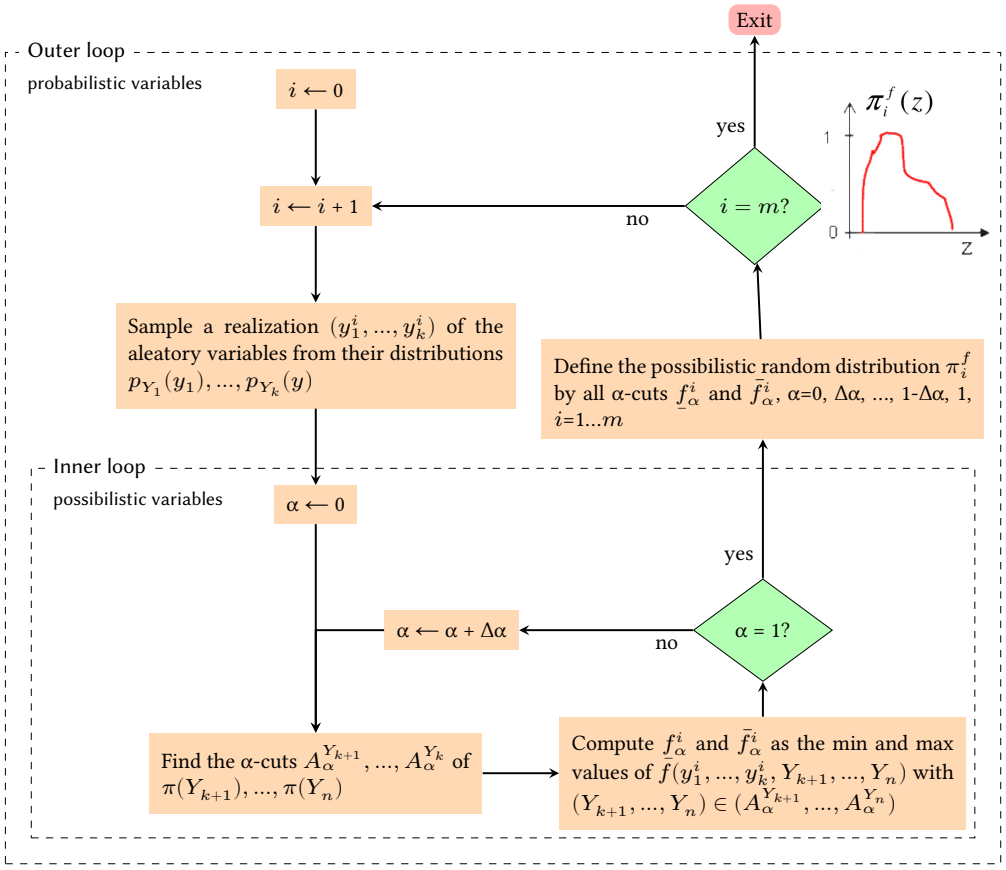


Figure 4.5 – *Scheme of the operative steps of the hybrid method: in evidence an example of fuzzy random realization* π_i^f *of* $f(Y)$

from $\mathrm{Pl}\left(f(Y) \in (-\infty, z]\right)$ and $\mathrm{Bel}\left(f(Y) \in (-\infty, z]\right)$, respectively **[Baudrit et al. 2006]**. On the other hand, $\mathrm{Bel}\left(f(Y) \in (-\infty, z]\right)$ and $\mathrm{Pl}\left(f(Y) \in (-\infty, z]\right)$ cannot convey any information on the prediction that $f(Y)$ lies within a given interval $[z_1, z_2]$, since neither $\mathrm{Bel}\left(f(Y) \in [z_1, z_2]\right)$ nor $\mathrm{Pl}\left(f(Y) \in [z_1, z_2]\right)$ can be expressed in terms of $\mathrm{Bel}\left(f(Y) \in (-\infty, z]\right)$ and $\mathrm{Pl}\left(f(Y) \in (-\infty, z]\right)$, respectively. A visual representation of this procedure is given in figure 4.6.

To prove equations (4.5) and (4.6), let us consider for simplicity a function $Z = f(Y_1, Y_2)$ of two variables Y_1, a discrete random variable with probability distribution $\{p_i^{Y_1}\}$, $i = 1...m$, and Y_2, a possibilistic variable with possibility distribution π^{Y_2}. Focal elements[1] for Y_1 are singletons $\{y_1^i; i = 1...m\}$ and the corresponding mass distribution is equal to $\left\{p_i^{Y_1}; i = 1...m\right\}$ because Y_1 is discrete. Focal elements for Y_2 corresponding to α-cuts are denoted $A_{\alpha_j}^{Y_2}$, $j = 1...q$ where q is the index of the α-cuts. The mass distribution associated to the focal elements (*i.e.*, the α-cuts, $A_{\alpha_j}^{Y_2}$), is denoted $v_j^{Y_2} = \alpha_j - \alpha_{j+1}, j = 1...q$ **[Baudrit et al. 2006]**.

Under the hybrid method, $Z = f(Y_1, Y_2)$, is a discrete random fuzzy subset: that is, m fuzzy random realizations π_i^f, $i = 1...m$, are obtained with probabilities $\{p_i^{Y_1}; i = 1...m\}$. Under the random set approach **[Baudrit and Dubois 2005]**, this random fuzzy set is interpreted as $m \times q$ focal elements (intervals) with mass distributions $p_i^{Y_1} \times v_j^{Y_2}$, $i = 1...m, j = 1...q$, and focal elements $A_{ij}^f = f(y_1^i, A_{\alpha_j})$.

The calculation of the plausibility $\mathrm{Pl}^f(A)$ of a generic set A of values of $f(Y_1, Y_2)$ reads as follows:

$$\mathrm{Pl}^f(A) = \sum_{(i,j), A \cap A_{ij}^f \neq 0} p_i^{Y_1} \times v_j^{Y_2} = \sum_{i=1...m} p_i^{Y_1} \times \sum_{\substack{j=1...q \\ A \cap A_{ij}^f \neq 0}} v_j^{Y_2} = \sum_{i=1...m} p_i^{Y_1} \times \mathrm{Pl}_i^f(A)$$

Since we have $A_{ij}^f \subseteq ... \subseteq A_{ih}^f \quad \forall j \geq h$, then $\mathrm{Pl}_i^f(A) = \Pi_i^f(A)$ and $\mathrm{Pl}^f(A) = \sum_{i=1...m} p_i^{Y_1} \times \Pi_i^f(A)$. These results still hold when several independent probabilistic variables are involved, whereas they do not directly apply with more than one possibilistic variable. Indeed, recall that fuzzy arithmetic presupposes total dependence between α-cuts.

Consider now two discrete probabilistic variables, Y_1, Y_2, encoded by their focal elements (singletons) $\{y_1^i; i = 1...m\}, \{y_2^j; j = 1...m\}$ and the mass distributions $\left\{p_i^{Y_1}; i = 1...m\right\}, \left\{p_j^{Y_2}; j = 1...m\right\}$ and two possibilistic variables, Y_3, Y_4, encoded as belief functions, by their focal elements $A_{\alpha_k}^{Y_3}, k = 1...q, A_{\alpha_l}^{Y_4}, l = 1...q$, and the mass distributions $v_k^{Y_3} = \alpha_k - \alpha_{k+1}, k = 1...q$, $v_l^{Y_4} = \alpha_l - \alpha_{l+1}, l = 1...q$. If independence between focal sets is assumed, the joint mass distribution, v_{ijkl}, associated to focal elements $A_{\mathrm{ijkl}}^f = f\left(\{y_1^i\}, \{y_2^j\}, \pi_{\alpha_k}^{Y_3}, \pi_{\alpha_l}^{Y_4}\right)$ of $f(Y_1, Y_2, Y_3, Y_4)$, is defined by $v_{\mathrm{ijkl}} = p_i^{Y_1} \times p_j^{Y_2} \times v_k^{Y_3} \times v_l^{Y_4} \quad \forall i, j, k, l$. It corresponds to applying a Monte Carlo method to all variables: in particular, for each possibility distribution, an α-cut (here $A_{\alpha_k}^{Y_3}$ and $A_{\alpha_l}^{Y_4}$) is independently selected, thus assuming independence of the focal elements pertaining to different variables. On the contrary, in the case of total dependence between the focal elements of the possibilistic variables, *i.e.*, when the same value α is selected for all possibilistic variables (like in the hybrid method adopted in the present work), the joint possibility distribution π^{Y_3, Y_4} is characterized by $\min(\pi^{Y_3}, \pi^{Y_4})$ which corresponds to nested Cartesian products of α-cuts. Letting $v_k^{Y_3, Y_4} = v_k^{Y_3} = v_k^{Y_4}$ be the mass associated to the Cartesian product $A_{\alpha_k}^{Y_3} \times A_{\alpha_k}^{Y_4}$, we obtain:

$$\forall i, j, k, l, k = l: \quad v_{ijkk} = p_i^{Y_1} \times p_j^{Y_2} \times v_k^{Y_3, Y_4} \tag{4.7}$$

$$\forall i, j, k, l, k \neq l: \quad v_{ijkl} = 0 \tag{4.8}$$

Hence, the estimate of the plausibility, for all measurable sets A, is defined as:

$$\mathrm{Pl}^f(A) = \sum_{A \cap A_{ijkk}^f \neq 0} p_i^{Y_1} \times p_j^{Y_2} \times v_k^{Y_3, Y_4} = \sum_{\substack{i=1...m \\ j=1...m}} p_i^{Y_1} \times p_j^{Y_2} \times \sum_{\substack{k=1...q \\ A \cap A_{\mathrm{ijkk}}^f \neq 0}} v_k^{Y_3, Y_4} = \sum_{\substack{i=1...m \\ j=1...m}} p_i^{Y_1} \times p_j^{Y_2} \times \mathrm{Pl}_{\mathrm{ij}}^f(A)$$

[1] $U \subseteq W$ is called a *focal element* of the belief function Bel if and only if $m(U) > 0$, where m is the mass function of Bel.

Since we have $A^f_{ijkk} \subseteq \ldots \subseteq A^f_{ijhh} \forall k \geq h$, then $\mathrm{Pl}^f_{ij}(A) = \Pi^f_{ij}(A)$ and $\mathrm{Pl}^f(A) = \sum_{i=1\ldots m, j=1\ldots m} p_i^{Y_1} \times p_j^{Y_2} \times \Pi^f_{ij}(A)$ where $\Pi^f_{ij}(A)$ are the possibility measures associated to the output possibility distributions π^f_{ij} obtained by the hybrid method.

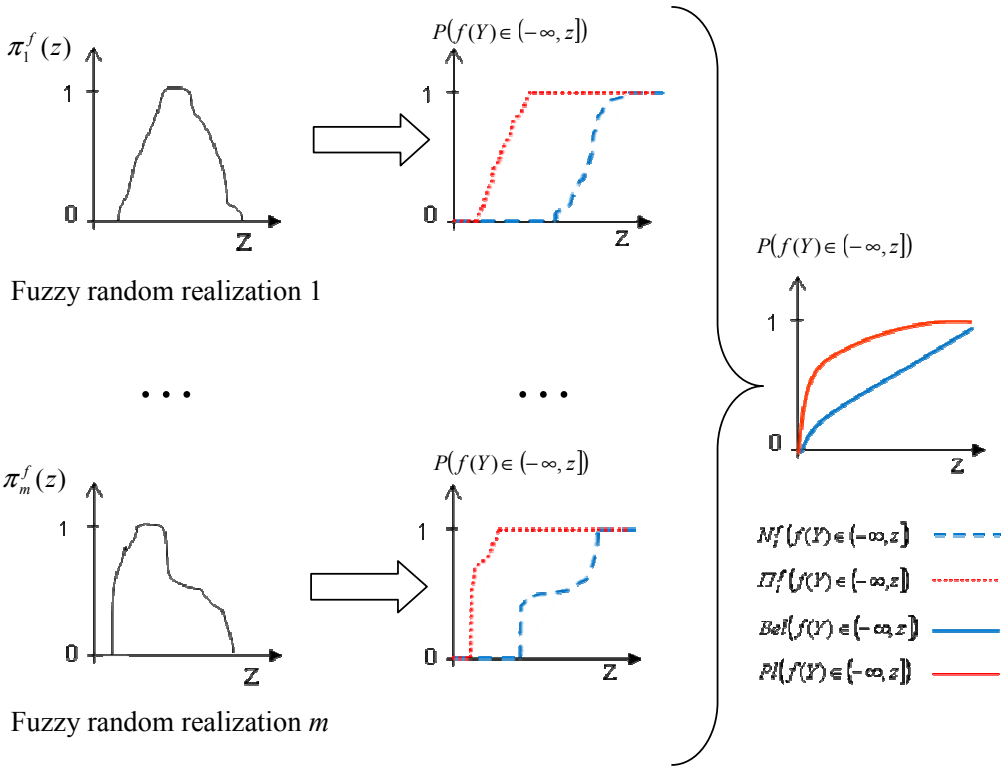


Figure 4.6 – *Left: fuzzy random realizations* $(\pi_1^f, \ldots, \pi_m^f)$ *of* $Z = f(Y)$*; middle: realizations of possibility and necessity measures,* $\Pi_i^f\left(f(Y) \in (-\infty, z]\right)$ *and* $N_i^f\left(f(Y) \in (-\infty, z]\right)$ *associated to* π_i^f*,* $i = 1\ldots m$*; right: belief and plausibility functions* $Bel\left(f(Y) \in (-\infty, z]\right)$ *and* $Pl\left(f(Y) \in (-\infty, z]\right)$ *obtained using equations (4.5) and (4.6).*

4.4 Application

The case study deals with the design of a protection dike in a residential area that is closely located to a river with potential risk of floods. As a mitigation and prevention measure, a dike has to be built to protect the area. Two issues of concern are: i) high construction and annual maintenance costs of the dike; ii) uncertainty in the natural phenomenon of flooding. Then, the different design options must be evaluated within a flooding risk analysis framework accounting for uncertainty **[Limbourg and de Rocquigny 2010]**.

In this section, the purely probabilistic approach (§ 4.3.1) and the Monte Carlo simulation and fuzzy interval analysis approach (§ 4.3.2) are applied to the case study. In particular, in § 4.4.1, the description of the model is given; in § 4.4.2, the model input variables and the representation of the associated uncertainties are illustrated; in § 4.4.3, the results of the uncertainty propagation are reported.

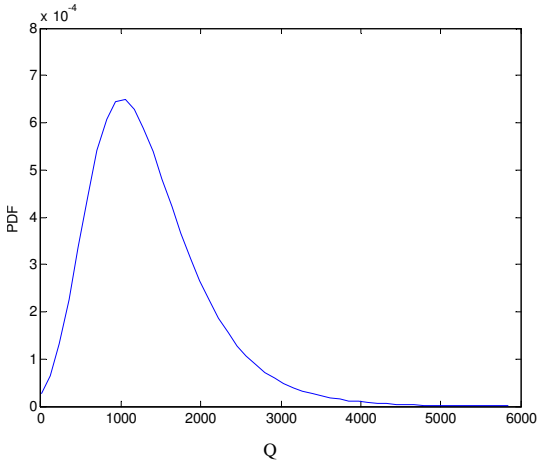


Figure 4.7 – *Gumbel probability distribution function for the maximal water flow Q [m³/s]*

4.4.1 The model

The maximal water level of the river (*i.e.*, the output variable of the model, Z_c) is given as a function of several parameters (*i.e.*, the input variables of the model), some of which are uncertain **[Limbourg and de Rocquigny 2010]**:

$$Z_c = Z_v + \left(\frac{Q}{K_s \times B \times \sqrt{(Z_m - Z_v)/L}} \right)^{3/5} \tag{4.9}$$

where:

- ▷ Q is the yearly maximal water discharge (m^3/s);
- ▷ Z_m and $Z_m Z_v$ are the riverbed levels (m asl) at the upstream and downstream part of the river under investigation, respectively;
- ▷ K_s is the Strickler friction coefficient;
- ▷ B and L are the width and length of the river part (m), respectively.

The input variables are classified as follows:

- ▷ *Constants:* $B = 300$ m, $L = 5000$ m.
- ▷ *Uncertain variables:* Q, Z_m, $Z_m Z_v$, K_s.

4.4.2 The input variables: physical description and representation of the associated uncertainty

A physical description of the input variables and the associated uncertainty representation is given in the following.

The *maximal water flow* Q is the variable with the largest amount of data available. A large set of water flow data for the river given is available to perform Bayesian inference (149 annual maximal flow values) on the parameters of the distribution. The Gumbel distribution $\mathrm{Gum}(q|\alpha, \beta)$ is a well-established probabilistic model for maximal flows:

$$\mathrm{Gum}(q|\alpha, \beta) = \frac{1}{\beta} \exp\left[-\exp\left(\frac{q - \alpha}{\beta} \right) \right] \exp\left[\frac{\alpha - q}{\beta} \right] \tag{4.10}$$

The Bayesian posterior estimates of the parameters of the distribution are $\alpha = 1014.0$ and $\beta = 565.4$ **[Pasanisi et al. 2009]**; the corresponding probability density function is shown in figure 4.7.

The uncertainties in the *upstream and downstream levels*, Z_m and $Z_m Z_v$ respectively, are quantified by a bivariate normal distribution $N(\mu, \sigma)$. Indeed, as the upstream and downstream sections are quite close it seems reasonable to model them as possibly dependent variables. A total of 29 pairs of data $(Z_m^{(i)}, Z_v^{(i)})$ have been used to perform Bayesian inference and setting the posterior distribution for μ and σ. The point values μ^* and σ^* used in this chapter for the

distribution parameters μ and σ are $\mu^* = [\mu^*_m, \mu^*_v] = [55.0, 50.2]$ and $\sigma^* = \left[\sigma^*_m, \sigma^*_{mv}, \sigma^*_{vm}, \sigma^*_v\right] = [0.46, 0.3388; 0.3388, 0.39]$: these are the mean values of the Bayesian posterior probability density functions of the parameters μ and σ of the probabilistic distributions of the uncertain variables Z_m and $Z_m Z_v$ **[Pasanisi et al. 2009]**. The resulting probability density functions are shown in figure 4.8.

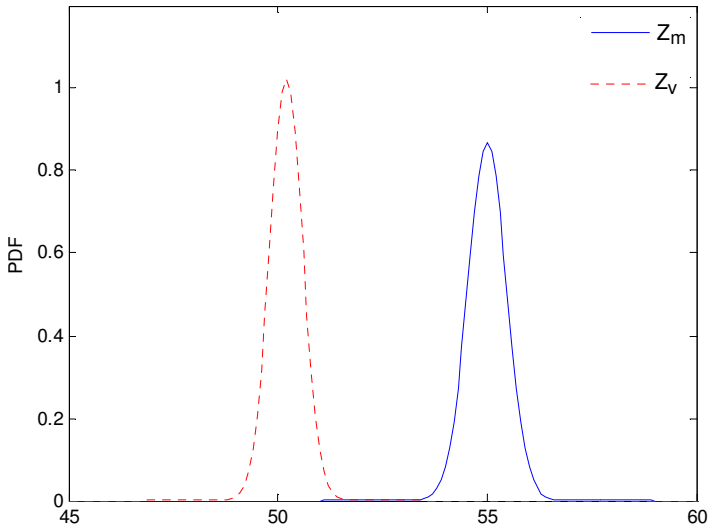


Figure 4.8 – *Normal probability distribution functions of the levels upstream Z_m (solid lines) and downstream Z_v (dashed lines)*

The *Strickler friction coefficient* K_s, an empirical coefficient which represents the surface roughness and sinuosity, is perhaps the most critical source of uncertainty. It is affected by epistemic uncertainty, since it is a simplification of a much more complex hydraulic model. In addition, assessing the uncertainty of K_s is difficult because, in practice, even if this coefficient is strongly related to the morphology of the river, it cannot be measured. As a consequence, data may only be retrieved through indirect calibration noised by significant observational uncertainty: this is reflected in only a very small series available of 5 data sets with ± 15% noise **[Limbourg and de Rocquigny 2010]**. The absolute physical limits of K_s are $[a, b] = [5, 60]$, but the real value is expected to vary in a smaller range.

In **[Pasanisi et al. 2009]**, this epistemic variable is treated within a probabilistic framework: it is considered that the probability distribution of K_s is normal with mean μ and standard deviation σ equal to 30 and 7.5, respectively. In this work, the epistemic uncertainty associated to K_s is represented by means of possibility distributions; the four methods described in the previous section are used to this aim. Note that for the method of § 4.2.1 (*i.e.*, triangular possibility distribution), the basis of the triangle is [5, 60] (*i.e.*, the absolute physical limits of K_s) and the most likely value is 30 (*i.e.*, the mean μ of the normal probability density function of K_s used in **[Pasanisi et al. 2009]**); for the methods of § 4.2.1 and § 4.2.1 (*i.e.*, Chebyshev inequality and probability-possibility transformations) the mean μ and the standard deviation σ used are 30 and 7.5 (this corresponds to the mean and the standard deviation of the probability density function of K_s used in **[Pasanisi et al. 2009]**).

The possibility distributions for K_s resulting from the application of the methods in § 4.2.1 are shown in figure 4.9. It is worth noting that the area lying under the possibility distribution is related to the imprecision in the knowledge of the possibilistic variable: the larger the area, the higher the imprecision.

It can be noticed that the larger areas are those underlying the possibility distributions built using the triangular function and the Chebyshev inequality. In fact, the information available to the analyst for building these two possibility distributions is quite scarce: in the first case, only the physical limits and the most likely value of the variable are known; in the second case, only the mean value and the standard deviation are considered. On the contrary, the smaller areas underlying the possibility distributions constructed by the transformation methods are explained by the larger amount of information available to the analyst concerning the epistemic variable of interest, *i.e.*, the probability distribution function itself **[Baraldi et al. 2012]**.

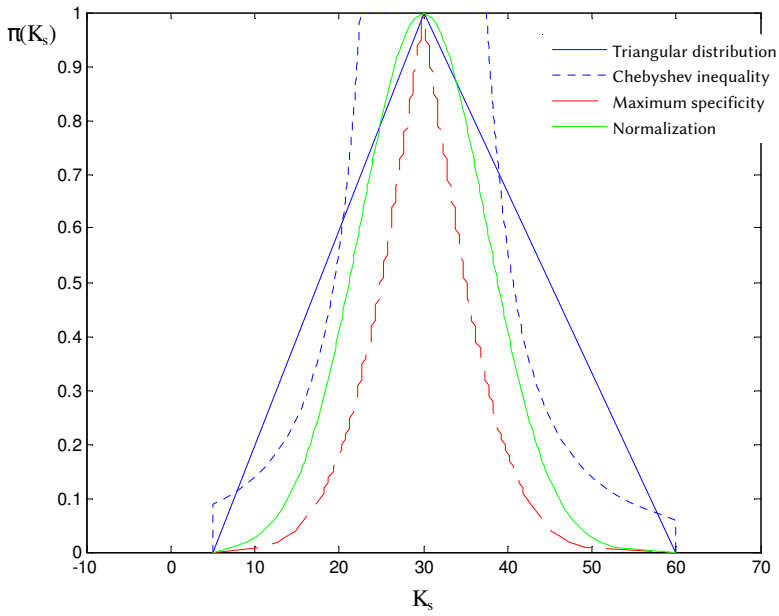


Figure 4.9 – *Comparison of the four different possibility distributions used for K_s: triangular function, Chebyshev inequality, principle of maximum specificity, normalization of the probability density* **[Baraldi et al. 2012]**.

For each of the four cases considered (*i.e.*, for each of the possibility distributions built), the hybrid approach of § 4.3.2 is run with m = 10 000 realizations of the probabilistic variables; for each realization of the probabilistic variables, 21 values of α (0, 0.05, 0.1, ..., 1) are considered to process the epistemic uncertainty associated with K_s. The results of the hybrid approach are compared to those obtained with a pure probabilistic approach with n = 10 000 samplings of the probabilistic variables: in this case, the probability distribution of K_s is considered normal with mean μ and standard deviation σ equal to 30 and 7.5, respectively, as in the reference paper **[Pasanisi et al. 2009]**.

To illustrate the procedure proposed in § 4.3.2, the uncertainty propagation in the computation of the maximal water level of the river $Z_c = Z_v + \left(\frac{Q}{K_s \times 300 \times \sqrt{(Z_m - Z_v)/5000}}\right)^{3/5}$, $Z_c = f(Q, Z_m, Z_v, K_s)$, is illustrated step by step **[Baraldi et al. 2012]**:

1. A number m = 10 000 of realizations of the probabilistic variables (Q^i, Z_m^i, Z_v^i) has been sampled from the corresponding probability density functions proposed in **[Pasanisi et al. 2009]** and shown in figure 4.7 and figure 4.8. Then, for each realization, steps 2. – 5. below have been performed.

2. With respect to the scarce information about the epistemic variable K_s, 21 values of α (0, 0.05, 0.1, ..., 1) have been considered. The α-cuts $[\underline{K}_s^{\alpha}, \overline{K}_s^{\alpha}]$ of the corresponding possibilistic distribution $\pi(K_s)$ have been found. For each α-cut, steps 3. – 5. below have been performed.

3. For the i^{th} realization of the probabilistic variables, Q^i, Z_m^i, Z_v^i, and the α-cut of the possibilistic variable, K_s, the smallest $\underline{f}_{\alpha}^{i}$ and largest $\overline{f}_{\alpha}^{i}$ values of f have been computed considering the fixed values Q^i, Z_m^i, Z_v^i and all values of K_s in its α-cut interval $[\underline{K}_s^{\alpha}, \overline{K}_s^{\alpha}]$. In this particular case, given the structure of f, $\underline{f}_{\alpha}^{i} = Z_v^i + \left(\frac{Q^i}{\overline{K}_s^{\alpha} \times 300 \times \sqrt{(Z_m^i - Z_v^i)/5000}}\right)^{3/5}$ and $\overline{f}_{\alpha}^{i} = Z_v^i + \left(\frac{Q^i}{\underline{K}_s^{\alpha} \times 300 \times \sqrt{(Z_m^i - Z_v^i)/5000}}\right)^{3/5}$.

4. The extreme values $\underline{f}_{\alpha}^{i}$ and largest $\overline{f}_{\alpha}^{i}$ found in 3. have been taken as the lower and upper limit of the α-cut of f for the i^{th} realization.

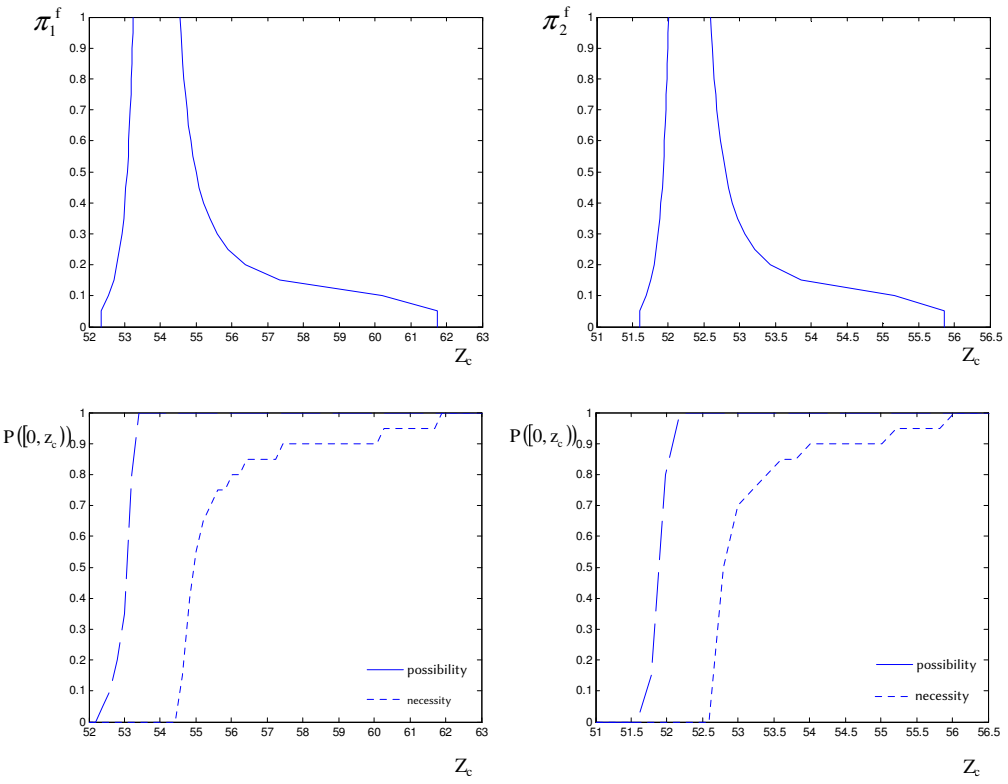


Figure 4.10 – *Top: two different fuzzy random realizations of the maximal water level of the river. Bottom: corresponding necessity and possibility measures.*

5. After having repeated steps 3. and 4. for all the 21 α-cuts found in step 2., the fuzzy random realization of f is constructed as the collection of its 21 α-cut intervals $\left[\underline{f}_{\alpha}^{i}, \overline{f}_{\alpha}^{i}\right]$.

6. Steps 2. – 5. have been repeated for each of the $m = 10\,000$ realizations Q^i, Z_m^i, Z_v^i, $i = 1\ldots10\,000$, producing m fuzzy random realizations of π_i^f. Figure 4.10 (top) shows two different examples of fuzzy random realizations of the maximal water level of the river Z_c.

Then, for all sets $A = [0, z_c)$, $z_c \in R^+$ the possibility and the necessity measures, $\Pi_i^f\left([0, z_c)\right)$ and $N_i^f\left([0, z_c)\right)$, are obtained from the corresponding possibility distributions $\pi_i^f(z_c)$, according to (4.3) and (4.4), respectively. Figure 4.10 (bottom) reports the possibility and necessity measures corresponding to the possibility distributions shown in figure 4.10 (top).

Finally, the $m = 10\,000$ possibility and necessity measures are combined to obtain the belief and plausibility measures by (4.5) and (4.6): $\text{Bel}\left([0, z_c)\right) = \sum_{i=1}^{m} \frac{1}{m} N_i^f\left([0, z_c)\right)$ and $\text{Pl}\left([0, z_c)\right) = \sum_{i=1}^{m} \frac{1}{m} \Pi_i^f\left([0, z_c)\right)$. These results are reported in § 4.4.3.

4.4.3 Results of the uncertainty propagation

In this section, the results of the uncertainty propagation are reported. The Monte Carlo simulation and fuzzy interval analysis and the pure probabilistic approaches are compared, then a sensitivity analysis of the results of the fuzzy interval analysis propagation is presented.

Comparison between the Monte Carlo simulation and fuzzy interval analysis approach with the pure probabilistic approach

Figures 4.11 to 4.14 show the comparison of the cumulative distribution functions of the maximal water level of the river (*i.e.*, the output variable of the model, Z_c) obtained by the probabilistic uncertainty propagation approach (solid lines) with the belief (lower curves) and plausibility (upper curves) functions obtained by the hybrid approach, where the possibility distributions for K_s are constructed using the methods of § 4.2.1.

It can be seen that **[Baraldi et al. 2012]**:

- The hybrid approach explicitly propagates the uncertainty by separating the contributions coming from the probabilistic and possibilistic variables; this separation is visible in the output distributions of the maximal water level of the river where the *separation* between the belief and plausibility functions reflects the imprecision in the knowledge of the possibilistic variable K_s.
- The *separation* between the belief and plausibility functions is larger for the cases in figure 4.11 and figure 4.12 (where the possibility distributions are those of built using the triangular function and Chebyschev inequality, respectively) with respect to those in figure 4.13 and figure 4.14 (where the possibility distributions are those built using the probability-possibility transformations); the larger gap between the belief and plausibility functions in figure 4.11 and figure 4.12 than in figure 4.13 and figure 4.14 is explained by the larger *area* contained under the corresponding possibility distribution functions (actually, the larger the area, the higher the imprecision in the knowledge of the possibilistic variable).
- The uncertainty in the output distribution of the pure probabilistic approach is given *only* by the *slope* of the cumulative distribution.
- As expected, the cumulative distribution of the maximal water level of the river obtained by the pure probabilistic method is within the belief and plausibility functions obtained by the hybrid approach.

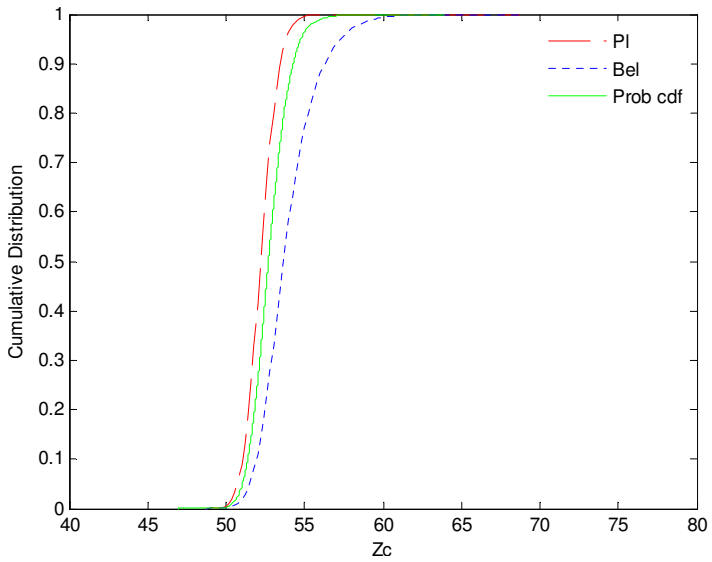


Figure 4.11 – *Comparison of the cumulative distribution functions of the maximal water level of the river Z_c obtained by the probabilistic uncertainty propagation approach (solid line) with the belief (lower dashed curve) and plausibility (upper dashed curve) functions obtained by the hybrid approach with the possibility distribution of K_s taken as a* ***triangular function*** *(see § 4.2.1 and figure 4.9)* **[Baraldi et al. 2012]**

The final goal of the uncertainty study is to determine i) the dike level necessary to guarantee a given flood return period or ii) the flood risk for a given dike level.

With respect to issue i) above, the quantity of interest that is mostly relevant to the decision-maker is the 99% quantile of Z_c, *i.e.*, $Z_c^{0.99}$, taken as the annual maximal flood level. This

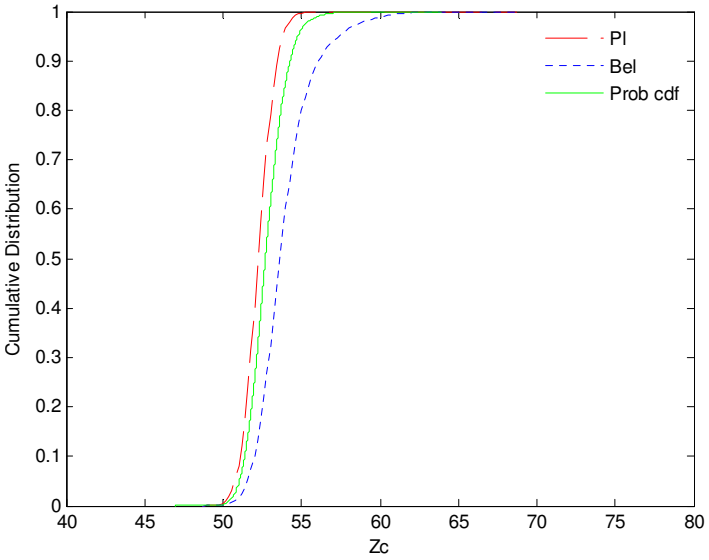


Figure 4.12 – *Comparison of the cumulative distribution functions of the maximal water level of the river Z_c obtained by the probabilistic uncertainty propagation approach (solid line) with the belief (lower dashed curve) and plausibility (upper dashed curve) functions obtained by the hybrid approach with the possibility distribution of K_s built using the* ***Chebyshev inequality*** *(see § 4.2.1 and figure 4.9)* [Baraldi et al. 2012]

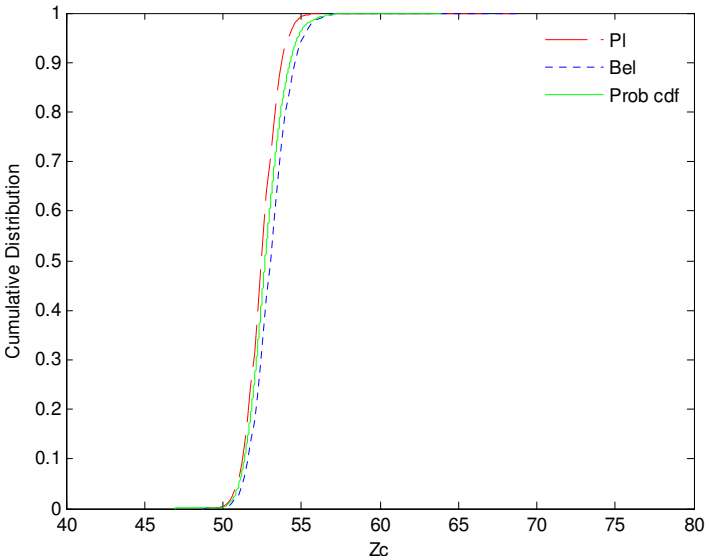


Figure 4.13 – *Comparison of the cumulative distribution functions of the maximal water level of the river Z_c obtained by the probabilistic uncertainty propagation approach (solid line) with the belief (lower dashed curve) and plausibility (upper dashed curve) functions obtained by the hybrid approach with the possibility distribution of K_s built on the transformation from probability to possibility distribution using the* ***principle of maximum specificity*** *(see § 4.2.1 and figure 4.9)* [Baraldi et al. 2012]

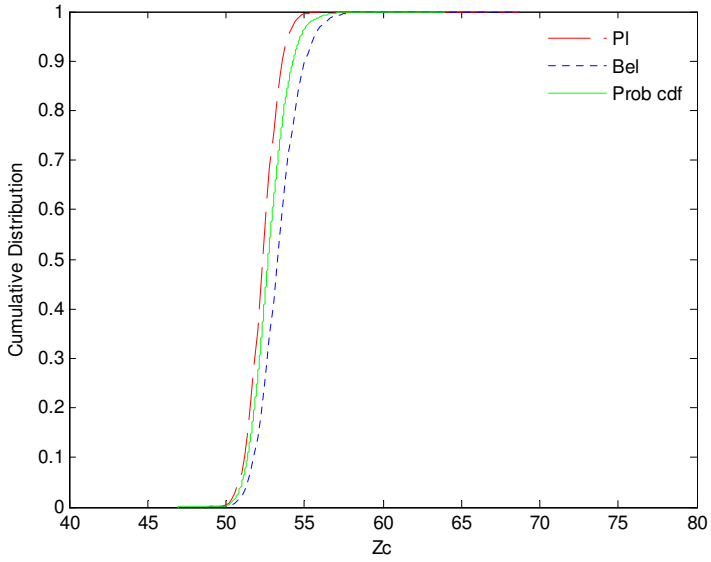


Figure 4.14 – *Comparison of the cumulative distribution functions of the maximal water level of the river Z_c obtained by the probabilistic uncertainty propagation approach (solid line) with the belief (lower dashed curve) and plausibility (upper dashed curve) functions obtained by the hybrid approach with the possibility distribution of K_s built on the transformation from probability to possibility distribution using the* ***normalization of probability density*** *(see § 4.2.1 and figure 4.9)* **[Baraldi et al. 2012]**

corresponds to the level of a "centennial" flood, the yearly maximal water level with a 100 year return period. With respect to issue ii) above, the quantity of interest that is mostly relevant to the decision-maker is the probability that the maximal water level of the river Z_c exceeds a given threshold z^*, *i.e.*, $P(Z_c > z^*)$; in the present document, $z^* = 55.5$ m as in **[Limbourg and de Rocquigny 2010]**. Table 4.1 reports the lower ($Z^{0.99}_{c,\text{lower}}$) and upper ($Z^{0.99}_{c,\text{upper}}$) 99$^{\text{th}}$ percentiles obtained from the two limiting cumulative distributions by using the four different possibility distributions proposed in § 4.2.1 (*i.e.*, triangular function, Chebyshev inequality, principle of maximum specificity and normalization of the probability density function) and the corresponding $\text{Bel}(Z_c > z^*)$ and $\text{Pl}(Z_c > z^*)$. In addition, as synthetic mathematical indicators of the imprecision in the knowledge of Z_c (*i.e.*, of the separation between the belief and plausibility functions), the percentage widths:

- ▷ $W_{Zc} = \frac{Z^{0.99}_{c,\text{upper}} - Z^{0.99}_{c,\text{lower}}}{Z^{0.99}_{c,\text{prob}}}$ of the interval $[Z^{0.99}_{c,\text{lower}}, Z^{0.99}_{c,\text{upper}}]$ with respect to the percentile $Z^{0.99}_{c,\text{prob}}$ obtained by the pure probabilistic approach
- ▷ $W^* = \frac{Pl(Z_c>z^*) - Bel(Z_c>z^*)}{P(Z_c>z^*)_{prob}}$ of the interval $\left[Bel(Z_c > z^*), Pl(Z_c > z^*)\right]$.

have been reported.

The numerical results in table 4.1 confirm the similarities between the cumulative distributions obtained by using the triangular function and the Chebyshev inequality for the possibilistic representation of the uncertainty on K_s, and between the cumulative distributions obtained by the two different transformations from probability to possibility distributions **[Baraldi et al. 2012]**.

Possibility distribution	$Z_c^{0.99}$ (pure probabilistic value = 56.10)		P[$Z_c \geq 55.5$] (pure probabilistic value = 0.0191)	
	$\left[Z_{c,\text{lower}}^{0.99}, Z_{c,\text{upper}}^{0.99}\right]$	W_{Zc} [%]	[Bel, Pl]	W^* [%]
Triangular function	[54.57, 59.29]	8	[0.0015, 0.1682]	873
Chebyshev inequality	[54.40, 60.00]	10	[0.0014, 0.1631]	847
Transformation probability to possibility (principle of maximum specificity)	[54.60, 56.69]	4	[0.0028, 0.0705]	355
Transformation probability to possibility (normalization)	[54.83, 55.99]	2	[0.0043, 0.0344]	157

Table 4.1 – *Lower and upper values of the Z_c percentiles and the threshold exceedance probability, and calculation of the indicator W about the width of the confidence interval* [Baraldi et al. 2012]

Sensitivity analysis with respect to the fuzzy interval analysis propagation

We have also analyzed the sensitivity of the output limiting cumulative functions to a different choice of the number of α-cuts considered for processing the epistemic uncertainty of the Strickler coefficient, K_s. The analysis was carried out considering the possibility distribution of K_s obtained by using the Chebyshev inequality. Three cases are taken into account: 5, 20, 100 α-cuts. The results of the analysis, reported in figure 4.15, show that in this case study a good trade-off between precision of the results and computation time[2] is achieved by using 20 α-cuts for the propagation of the uncertainty described by the possibility distributions. Reducing the number of α-cuts leads to imprecise estimation of the quantities of interest, whereas an higher number of α-cuts causes a significant increase of the computational time.

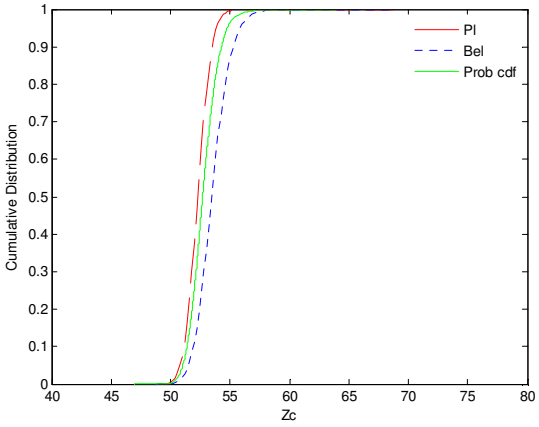
4.5 Conclusions

In this report, an 'hybrid' computational framework has been presented, which allows the joint propagation of probabilistic and possibilistic uncertainty representations. An application to a flood risk model has been illustrated as a realistic benchmark for uncertainty modeling. Aleatory and epistemic uncertainties have been kept separate in the model, *i.e.*, some of the variables are purely probabilistic (aleatory uncertainty) and some are purely possibilistic (epistemic uncertainty).

The following analyses have been carried out:

1. A comparison has been performed between the 'hybrid' and the 'pure probabilistic' approach, highlighting that:
   - ▷ The uncertainty in the output distribution of the pure probabilistic approach is given *only* by the *slope* of the cumulative distribution.
   - ▷ The hybrid approach explicitly propagates the uncertainty by separating the contributions coming from the probabilistic and possibilistic variables.
   - ▷ The larger gap between the belief and plausibility functions is explained by the larger *area* contained under the corresponding possibility distribution functions.
   - ▷ As expected, the cumulative distribution of the model output obtained by the pure probabilistic method is within the belief and plausibility functions obtained by the hybrid approach.
2. Four methods for constructing the possibility distributions of the variables subject to epistemic uncertainty have been compared, showing that:
   - ▷ The choice of the possibility distribution depends on the information available about the variable: when the physical limits and the most likely value are available, a triangular possibility distribution can be constructed; when the mean and the standard deviation can be computed, for instance using empirical data, the Chebyshev

[2] All reported computation times are on a Pentium 4 CPU operating at 3GHz.



5 α-cuts

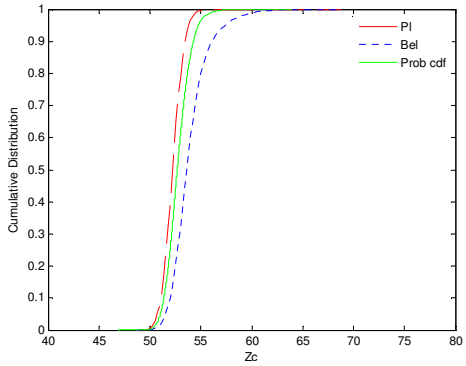
Computation time: 87 seconds

$Z_c^{0.99}$:

- ▷ $\left[Z_{c,\text{lower}}^{0.99}, Z_{c,\text{upper}}^{0.99}\right]$ = [54.41, 57.04]
- ▷ W_{Z_c}: 5%

P[Z ≥ 55.5]:

- ▷ [Bel, Pl] = [0.0012, 0.0828]
- ▷ W^* = 427%



20 α-cuts

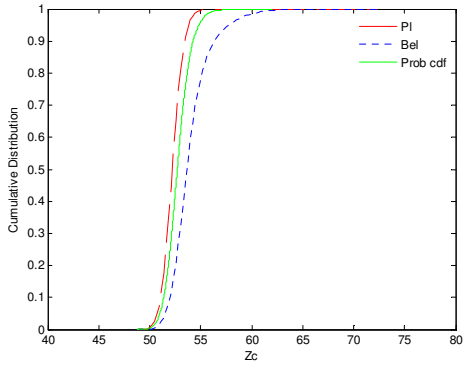
Computation time: 133 seconds

$Z_c^{0.99}$:

- ▷ $\left[Z_{c,\text{lower}}^{0.99}, Z_{c,\text{upper}}^{0.99}\right]$ = [54.40, 60.00]
- ▷ W_{Z_c}: 10%

P[Z ≥ 55.5]:

- ▷ [Bel, Pl] = [0.0014, 0.1631]
- ▷ W^* = 847%



100 α-cuts

Computation time: 321 seconds

$Z_c^{0.99}$:

- ▷ $\left[Z_{c,\text{lower}}^{0.99}, Z_{c,\text{upper}}^{0.99}\right]$ = [54.47, 60.66]
- ▷ W_{Z_c}: 11%

P[Z ≥ 55.5]:

- ▷ [Bel, Pl] = [0.0011, 0.1669]
- ▷ W^* = 868%

Figure 4.15 – *On the left: comparison of the cumulative distribution functions of the maximal water level of the river Z_c obtained by the probabilistic uncertainty propagation approach (solid line) with the belief (lower dashed curve) and plausibility (upper dashed curve) functions obtained by the hybrid approach with a different number of α-cuts. On the right: computation time, lower and upper values of Z_c percentiles, threshold exceedance probability and the respective percentage width W of the intervals.*

inequality can be used; when a probability distribution is available, the methods for transforming probability into possibility distributions can be employed.

- ▷ There are similarities between the results obtained by using:
  (a) the triangular function and the Chebyshev inequality;
  (b) the two transformations from probability to possibility distributions (*i.e.*, those based on the principle of maximum specificity and on the normalization of the probability density function).

These similarities are explained by the same 'uncertainty content' borne by the corresponding possibility distributions (as demonstrated by the similar area limited by the possibility distribution functions).

3. An analysis of the sensitivity of the results to the number of α-cuts has been carried out, showing that:
   - ▷ a decrease in the number of α-cuts leads to imprecise estimates of the belief and plausibility functions of the output;
   - ▷ on the other hand, increasing the number of α-cuts causes a remarkable increase in the computational time.

   Thus, the choice of the number of α-cuts is driven by the trade-off between estimation accuracy and computational cost.

5

Conclusions

In this document, two procedures have been presented for ranking system components in order of importance when in presence of epistemic uncertainties affecting the components' reliability and availability parameters. One procedure allows accounting for epistemic uncertainties described by probability distributions (§ 1.3.2) **[Baraldi et al. 2009b]**; the other one allows handling epistemic uncertainties described by possibility distributions (§ 1.4) **[Baraldi et al. 2009a]**. In both methods, the ranking procedure is based on a pairwise comparison criterion that permits to establish a relation order between the uncertain importance measures (IMs) of two components.

Generally speaking, the main pitfall of the ranking approaches based on pairwise comparisons lies in the combinatorial explosion of the number of pairs of elements to be compared, when the systems are made up of a large number of components. The Quicksort algorithm has been used to partially overcome this drawback, since it allows sorting a large set of components on the basis of their importance, with a relatively small number of comparisons.

The application of the procedures to two case studies has shown that **[Baraldi et al. 2009b,a]**:

- ▷ Accounting for uncertainties in the computation of IMs is relevant: the ranking of the components' importances obtained neglecting the uncertainties can be different from that obtained by considering them.
- ▷ Compared to other approaches proposed in the literature **[Modarres 2006]**, the procedures presented in this document seem to overcome some limitations by a more satisfactory definition of the exceedance measures and a greater robustness of the final rank with respect to the choice of the pivot element in the sorting algorithm.
- ▷ The final ranking may depend on the investigation framework used to carry out the analysis, which is mainly established on the basis of the quality and quantity of available data. In general, the probabilistic representation and propagation of the uncertainty allows a more refined final ranking to be obtained, but it calls for a larger amount of available data and more accurate information (which may be lacking in real industrial applications).
- ▷ The final ranking may also depend on the choice of the pivot and on the initial arrangement of the components. In this respect, the execution of the sorting algorithm for different settings of these two parameters, would give more confidence on the final result.

A final remark is in order with respect to the cases where the epistemic uncertainties are represented by possibility distributions. The proposed procedures have been applied in both case studies to rank the components on the basis of their Birnbaum IM (§ 2.2 and § 3.2); this has led to a computational simplification. In fact, the simple rules of the fuzzy addition, subtraction and multiplication suffice to propagate the uncertainty from the basic events to the Birnbaum IMs of the components, and make very simple the search for their maxima and minima when the probabilities of the basic events range in a given α-cut (equations (1.9) and (1.10)) at step 1.2 of the procedure of § 1.4). The computation of other IMs (*e.g.*, the RRW, RAW, FV) makes it necessary to introduce also the division operation, which renders the search more difficult. Obviously, the larger the number of basic events, the larger the dimension of the space in which the maxima and minima must be searched for, and the larger the required computational effort. In case of complex systems, the application of optimization techniques such as Genetic Algorithms may be beneficial to reduce computation times. This is an open issue, which will be investigated in future work.

A

PDFs and CDFs of the difference between two uniform random variables

Given a generic uniform random variable $x \approx U(a, b)$, its moment generating function (mgf) is given by:

$$\phi(s) = \frac{e^{sb} - e^{sa}}{s(b-a)}$$

The random variable $r = I_A - I_B$ is the convolution of two uniformly distributed random variables and in particular:

$$I_A \approx U(a_{I_A}, b_{I_A})$$

$$-I_B \approx U(-b_{I_B}, -a_{I_B})$$

The mgf of r is given by:

$$\phi_r(s) = \frac{e^{sb_{I_A}} - e^{sa_{I_A}}}{s(b_{I_A} - a_{I_A})} \cdot \frac{e^{s(-a_{I_B})} - e^{s(-b_{I_B})}}{s(b_{I_B} - a_{I_B})}$$

$$\phi_r(s) = \frac{e^{s(b_{I_A} - a_{I_B})} - e^{s(a_{I_A} - a_{I_B})} - e^{s(b_{I_A} - b_{I_B})} + e^{s(a_{I_A}) - b_{I_B})}}{s^2(b_{I_A} - a_{I_A}) \cdot (b_{I_B} - a_{I_B})}$$

As for the inverse transformation, it could be noted that the mgf of r can be regarded as the algebraic sum of functions which are linearly increasing/decreasing with the same slope. So the pdf and cdf of r are given by:

$$f_r(r) = \begin{cases} \frac{r + b_{I_B} - a_{I_A}}{(b_{I_A} - a_{I_A}) \cdot (b_{I_B} - a_{I_B})} & a_{I_A} - b_{I_B} \le r \le b_{I_A} - b_{I_B} \\ \frac{1}{(b_{I_B} - a_{I_B})} & b_{I_A} - b_{I_B} \le r \le a_{I_A} - a_{I_B} \\ \frac{b_{I_A} - a_{I_B} - r}{(b_{I_A} - a_{I_A}) \cdot (b_{I_B} - a_{I_B})} & a_{I_A} - a_{I_B} \le r \le b_{I_A} - a_{I_B} \end{cases}$$

$$F_r(r) = \begin{cases} \frac{(b_{I_B} - a_{I_A} + r)^2}{2(b_{I_A} - a_{I_A}) \cdot (b_{I_B} - a_{I_B})} & a_{I_A} - b_{I_B} \le r \le b_{I_A} - b_{I_B} \\ 1 - \frac{(a_{I_A} + b_{I_A} - 2a_{I_B} - 2r)}{2(b_{I_B} - a_{I_B})} & b_{I_A} - b_{I_B} \le r \le a_{I_A} - a_{I_B} \\ 1 - \frac{(b_{I_A} - a_{I_B} - r)^2}{2(b_{I_A} - a_{I_A}) \cdot (b_{I_B} - a_{I_B})} & a_{I_A} - a_{I_B} \le r \le b_{I_A} - a_{I_B} \end{cases}$$

B

The Quicksort algorithm

A common method of simplification of a complex problem is to divide the problem into sub-problems of the same type: this technique in computer programming is called "divide and conquer". The quicksort algorithm applies such technique and sorts groups of elements by dividing their list (ordered array) into two sub-lists. In its simpler version the steps are:

- ▷ List the elements in an array ordered according to a given size parameter.
- ▷ Pick an element, called a pivot, from the list.
- ▷ Reorder the list so that all elements which are smaller than the pivot come before the pivot and all elements larger than the pivot come after it (equal values can go either way).
- ▷ After this partitioning, the pivot is in its final position. This is called the partition operation.
- ▷ Recursively sort the sub-list of smaller elements and the sub-list of larger elements, following the steps above.

The base case of the recursion (*i.e.* the stopping condition where the sorting function will not call itself anymore) are lists of size zero or one. In pseudocode, the algorithm can be described as follows:

```
function quicksort(array)
begin
    var list less, greater, equal
    if length(array) ≤ 1 then
        return array
    end
    pivot := the middle element of array
    put pivot into equal
    foreach x in array do
        if x < pivot then
            append x to less
        else if x > pivot then
            append x to greater
        else
            append x to equal
        end
    end
    return concatenate(quicksort(less), equal, quicksort(greater))
end
```

Notice that the elements are examined by comparing them to other elements: this makes quicksort a comparison sort algorithm. In general, the "equal" list is not defined, being equal values treated indifferently either as smaller or larger than the pivot. Since for the purpose of the present work, it is important to identify equalities in rank orders, a specific equal list has been added in the code.

The disadvantage of the simple version above is that it requires a lot of storage space. There exists a more sophisticated version which uses an in-place partition algorithm **[Knuth 1997]**,

which allows achieving the complete sort using a reduced memory space. The pseudocode is:

```
procedure quicksort(array, left, right)
begin
    if right < left then
        select a pivot index (e.g. pivotIndex ← left)
    end
    pivotNewIndex ← partition(array, left, right, pivotIndex)
    quicksort(array, left, pivotNewIndex - 1)
    quicksort(array, pivotNewIndex + 1, right)
end

function partition(array, left, right, pivotIndex)
begin
    pivotValue ← array[pivotIndex]
    swap array[pivotIndex] and array[right]
    storeIndex ← left
    for i from left to right - 1 do
        if array[i] ≤ pivotValue then
            swap array[i] and array[storeIndex]
            storeIndex ← storeIndex + 1
        end
        swap array[storeIndex] and array[right] // move pivot to its final place
    end
    return storeIndex
end
```

The algorithm partitions the portion of the array between indexes `left` and `right`, inclusively, by moving to the beginning of the subarray all elements smaller than or equal to a `pivotIndex`, leaving all the larger elements after. In the process, the final position for the pivot element is also found and temporarily moved to the end of the subarray, so that it does not interfere to the successive moves. Because only exchanges of positions are applied, the final list has the same elements as the original list. Notice that an element may be exchanged multiple times before reaching its final place. This kind of algorithm might be useful when a very large number of components has to be sorted.

C

Sorting algorithm proposed in [Modarres 2006]

This procedure follows the same steps 1 and 2 of the procedure in § 1.3.2, whereas it differs in the steps 3 and 4, which are as follows:

3. Find the probability that each component $i = 1...N$ occupies a specific position in the ranking. This is achieved by repeating for $v = 1...M$, the following Monte Carlo sampling:
   3.1. Sample a realization of the components' failure rates $\lambda_1^v, ..., \lambda_N^v$.
   3.2. Find the v-th importance measures relative to the failure rates of step 3.1.
   3.3. Rank the components' importance measures.
   3.4. The probability $P(R_i)$ that component i is in the rank position $R_i = 1, 2, ..., N$ is given by the ratio between the number of simulations with component i resulting in position R_i and the number of samples M.
4. To rank the component:
   4.1. List the components in the rank order found in step 1.
   4.2. Choose the most important component as pivot p, *i.e.* the component with largest probability of being the most important.
   4.3. Compute the measure of exceedance r_{pj}^* between the components p and j with $j = p + 1, p + 2$:

$$r_{pj}^* = P(R_p \geq R_j) = \sum_{R_p=1}^{n} p(R_i) \sum_{R_j=1}^{R_p} p(R_j)$$

   where R_p = rank of p and R_j = rank of j.
   4.4. If $r_{pj}^* > T_u$, then leave component p in the current position; else, if $T_l < r_{pj}^* < T_u$ then put the component j in position R_p; otherwise, if $r_{pj}^* < T_l$ swap the rank orders of components p and j.
   4.5. $p \leftarrow p + 1$, repeat steps 4.1–4.3 until $p = N$.

Bibliography

Anoop, M. B. and Rao, K. B. (2008). Determination of bounds on failure probability in the presence of hybrid uncertainties. *Sadhana*, 33(6):753–765. Available at http://www.ias.ac.in/sadhana/Pdf2008Dec/753.pdf.

Apostolakis, G. E. (1990). The concept of probability in safety assessment of technological systems. *Science*, 250(4986):1359–1364. DOI: 10.1126/science.2255906.

Apostolakis, G. E. and Kaplan, S. (1981). Pitfalls in risk calculations. *Reliability Engineering*, 2(2):135–145. DOI: 10.1016/0143-8174(81)90019-6.

Aven, T. (2010). On the need for restricting the probabilistic analysis in risk assessments to variability. *Risk Analysis*, 30(3):354–360. DOI: 10.1111/j.1539-6924.2009.01314.x.

Aven, T. (2011). On some recent definitions and analysis frameworks for risk, vulnerability, and resilience. *Risk Analysis*, 31(4):515–522. DOI: 10.1111/j.1539-6924.2010.01528.x.

Aven, T. and Steen, R. (2010). The concept of ignorance in a risk assessment and risk management context. *Reliability Engineering & System Safety*, 95(11):1117–1122. DOI: 10.1016/j.ress.2010.05.006.

Aven, T. and Zio, E. (2011). Some considerations on the treatment of uncertainties in risk assessment for practical decision making. *Reliability Engineering & System Safety*, 96(1):64–74. DOI: 10.1016/j.ress.2010.06.001.

Baraldi, P., Compare, M., and Zio, E. (2009a). Importance measures in presence of uncertainties. *Reliability and Risk Analysis: Theory and Applications*, 2(2):30–41. Available at http://gnedenko-forum.org/Journal/2009/022009/RATA_2_2009-03.pdf.

Baraldi, P., Compare, M., and Zio, E. (2009b). A method for ranking components importance in presence of epistemic uncertainties. *Journal of Loss Prevention in the Process Industries*, 22(5):582–592. DOI: 10.1016/j.jlp.2009.02.013.

Baraldi, P., Pedroni, N., Zio, E., Ferrario, E., Pasanisi, A., and Couplet, M. (2012). Monte Carlo and fuzzy interval propagation of hybrid uncertainties on a risk model for the design of a flood protection dike. In Bérenguer, C., Grall, A., and Soares, C. G., Ed., *Advances in Safety, Reliability and Risk Management — Proceedings of ESREL'2011*, Troyes, France. Available at http://hal.archives-ouvertes.fr/docs/00/65/80/77/PDF/anno_2011_5.pdf.

Baraldi, P. and Zio, E. (2008). A combined Monte Carlo and possibilistic approach to uncertainty propagation in event tree analysis. *Risk Analysis*, 28(5):1309–1325. DOI: 10.1111/j.1539-6924.2008.01085.x.

Baraldi, P. and Zio, E. (2010). A comparison between probabilistic and Dempster-Shafer theory approaches to model uncertainty analysis in the performance assessment of radioactive waste repositories. *Risk Analysis*, 30(7):1139–1156. DOI: 10.1111/j.1539-6924.2010.01416.x.

Baudrit, C. and Dubois, D. (2005). Comparing methods for joint objective and subjective uncertainty propagation with an example in a risk assessment. In Cozman, F. G., Nau, R., and Seidenfeld, T., Ed., *Proc. Fourth International Symposium on Imprecise Probabilities and their Applications (ISIPTA'05)*, Pittsburgh, PA, USA.

Baudrit, C. and Dubois, D. (2006). Practical representations of incomplete probabilistic knowledge. *Computational Statistics & Data Analysis*, 51(1):86–108. DOI: 10.1016/j.csda.2006.02.009.

Baudrit, C., Dubois, D., and Guyonnet, D. (2006). Joint propagation and exploitation of probabilistic and possibilistic information in risk assessment. *IEEE Transactions on Fuzzy Systems*, 14(5):593–608. DOI: 10.1109/TFUZZ.2006.876720.

Baudrit, C., Dubois, D., and Perrot, N. (2008). Representing parametric probabilistic models tainted with imprecision. *Fuzzy Sets and Systems*, 159(15):1913–1928. DOI: 10.1016/j.fss.2008.02.013.

Birnbaum, Z. W. (1969). Chapter *On the importance of different components in a multi component system* in *Multivariate analysis II* (Krishnaiah, P. R., Ed.), pages 581–592. Academic Press, New York. ISBN: 978-0124266520.

Bojadziev, G. and Bojadziev, M. (1995). *Fuzzy Sets, Fuzzy Logic, Applications*. World Scientific, Singapore. 283 pages.

Borgonovo, E. (2006). Measuring uncertainty importance: investigation and comparison of alternative approaches. *Risk Analysis*, 26(5):1349–1361. DOI: 10.1111/j.1539-6924.2006.00806.x.

Darbra, R. M., Eljarrat, E., and Barceló, D. (2008). How to measure uncertainties in environmental risk assessment. *Trends in Analytical Chemistry*, 27(4):377–385. DOI: 10.1016/j.trac.2008.02.005.

Dubois, D. (2006). Possibility theory and statistical reasoning. *Computational Statistics and Data Analysis*, 51:47–69. DOI: 10.1016/j.csda.2006.04.015.

Dubois, D. and Prade, H. (1980). *Fuzzy Sets and Systems: Theory and Applications*. Academic Press, New York, NY. ISBN: 0-12-222750-6, 393 pages.

Dubois, D. and Prade, H. (1988a). *Possibility theory: An Approach to Computerized Processing of Uncertainty*. Plenum Press, New York. ISBN: 978-0306425202, 280 pages.

Dubois, D. and Prade, H. (1988b). *Théorie des possibilités: application à la représentation des connaissances en informatique*. Masson, Paris. ISBN: 978-2225805790.

Dubois, D., Prade, H., and Sandri, S. (1993). Chapter *On possibility/probability transformations* in *Fuzzy Logic: State of the Art* (Lowen, R. and Roubens, M., Ed.), pages 103–112. Kluwer Academic Publishers, Dordrecht.

Dubois, D., Prade, H., and Smets, P. (1996). Representing partial ignorance. *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, 26(3):361–377. DOI: 10.1109/3468.487961.

EPA (2009). Guidance on the development, evaluation, and application of environmental models. Technical report EPA/100/K-09/003, U.S. Environmental Protection Agency, Council for Regulatory Environmental Modeling, Washington, DC, USA. Available at http://www.epa.gov/crem/library/cred_guidance_0309.pdf.

Fairman, R., Mead, C. D., Williams, W. P., and EEA (1998). Environmental risk assessment — approaches, experiences and information sources. Technical report 4/1998, European Environment Agency. Available at http://www.eea.europa.eu/publications/GH-07-97-595-EN-C2.

Ferson, S., den Brink, P. V., Estes, T. L., Gallagher, K., O'Connor, R., and Verdonck, F. (2010). Chapter *Bounding uncertainty analyses* in *Application of uncertainty analysis to ecological risks of pesticides* (Warren-Hicks, W. J. and Hart, A., Ed.), pages 89–122. CRC Press. ISBN: 978-1439807347.

Ferson, S. and Ginzburg, L. R. (1996). Different methods are needed to propagate ignorance and variability. *Reliability Engineering & System Safety*, 54(2):133–144. DOI: 10.1016/S0951-8320(96)00071-3.

Ferson, S. and Hajagos, J. G. (2004). Arithmetic with uncertain numbers: rigorous and (often) best possible answers. *Reliability Engineering & System Safety*, 85:135–152. DOI: 10.1016/j.ress.2004.03.008.

Ferson, S., Kreinovich, V., Ginzburg, L. R., Sentz, K., and Myers, D. S. (2003). Constructing probability boxes and Dempster-Shafer structures. Technical report, Sandia National Laboratories, Albuquerque, New Mexico. SAND2002-4015. Available at http://www.ramas.com/unabridged.zip.

Ferson, S., Kreinovich, V., Hajagos, J. G., Oberkampf, W. L., and Ginzburg, L. R. (2007). Experimental uncertainty estimation and statistics for data having interval uncertainty. Technical report, Sandia National Laboratories, Setauket, New York. SAND2007-0939. Available at http://www.ramas.com/intstats.pdf.

Ferson, S., Nelsen, R. B., Hajagos, J. G., Berleant, D., Zhang, J., Tucker, W. T., Ginzburg, L. R., and Oberkampf, W. L. (2004). Dependence in probabilistic modeling, Dempster-Shafer theory, and probability bounds analysis. Technical report, Sandia National Laboratories, Albuquerque, New Mexico. Technical Report SAND2004-3072.

Ferson, S. and Tucker, W. T. (2006). Sensitivity in risk analyses with uncertain numbers. Technical report, Sandia National Laboratories, Setauket, New York. Technical Report SAND2006-2801. Available at http://prod.sandia.gov/techlib/access-control.cgi/2006/062801.pdf.

Flage, R., Aven, T., and Zio, E. (2009). Alternative representations of uncertainty in reliability and risk analysis — review and discussion. In Martorell, S., Guedes, C. S., and Barnett, J., Ed., *Proceedings of the European safety and reliability conference 2009 (ESREL'2009)*, pages 2081–2091, London. CRC Press.

Flage, R., Baraldi, P., Ameruso, F., Zio, E., and Aven, T. (2010a). Handling epistemic uncertainties in fault tree analysis by probabilistic and possibilistic approaches. In Bris, R., Guedes, S. C., and Martorell, S., Ed., *Reliability, Risk and Safety: Theory and Applications. Supplement proceedings of the European Safety and Reliability Conference 2009 (ESREL'2009)*, Prague, Czech Republic.

Flage, R., Baraldi, P., Zio, E., and Aven, T. (2010b). Possibility-probability transformation in comparing different approaches to the treatment of epistemic uncertainties in a fault tree analysis. In *Proceedings of the European safety and reliability conference (ESREL'2010)*, Rhodes, Greece. CD-ROM.

Gil, M. A. (2001). Special issue on fuzzy random variables. *Information Sciences*, 133(1-2):1–2. DOI: 10.1016/S0020-0255(01)00072-X.

Halkidis, I. N., Tzimopoulos, D., Evangelides, H., and Sakellarioy-Makrantonaki, M. (2009). Soil water management problem using fuzzy arithmetic. *Global Nest Journal*, 11(4):556–565.

Helton, J. C., Johnson, J. D., Oberkampf, W. L., and Sallaberry, C. J. (2008). Representation of analysis results involving aleatory and epistemic uncertainty. Technical report, Sandia National Laboratories, Albuquerque, New Mexico. Technical Report SAND2008-4379.

Helton, J. C., Johnson, J. D., Oberkampf, W. L., and Storlie, C. B. (2007). A sampling-based computational strategy for the representation of epistemic uncertainty in model predictions with evidence theory. *Computer Methods in Applied Mechanics and Engineering*, 196(37-40):3980–3998. DOI: 10.1016/j.cma.2006.10.049.

Helton, J. C. and Oberkampf, W. L. (2004). An exploration of alternative approaches to the representation of uncertainty in model predictions. *Reliability Engineering & System Safety*, 85(1):39–71. Special issue on Alternative Representations of Epistemic Uncertainty. DOI: 10.1016/j.ress.2004.03.025.

Hoare, C. A. R. (1962). Quicksort. *Computer Journal*, 5(1):10–15. DOI: 10.1093/comjnl/5.1.10.

Huanga, D., Chenb, T., and Wang, M.-J. J. (2001). A fuzzy set approach for event tree analysis. *Fuzzy Sets and Systems*, 118(1):153–165. DOI: 10.1016/S0165-0114(98)00288-7.

Kalos, M. H. and Whitlock, P. A. (1986). *Monte Carlo methods. Volume I: Basics*. John Wiley & Sons, New York. ISBN: 978-0471898399, 186 pages.

Kendall, M. G. and Stuart, A. (1977). *The Advanced Theory of Statistics*. Griffin and Co., London, UK. ISBN: 978-0028478203.

Kentel, E. and Aral, M. M. (2004). Probabilistic-fuzzy health risk modeling. *Stochastic Environmental Research and Risk Assessment*, 18:324–338. DOI: 10.1007/s00477-004-0187-3.

Kentel, E. and Aral, M. M. (2005). 2D Monte Carlo versus 2D fuzzy Monte Carlo health risk assessment. *Stochastic Environmental Research and Risk Assessment*, 19(1):86–95. DOI: 10.1007/s00477-004-0209-1.

Kentel, E. and Aral, M. M. (2007). Risk tolerance measure for decision-making in fuzzy analysis: a health risk assessment perspective. *Stochastic Environmental Research and Risk Assessment*, 21(4):405–417. DOI: 10.1007/s00477-006-0073-2.

Klir, G. J. and Yuan, B. (1995). *Fuzzy Sets and Fuzzy Logic: Theory and Applications*. Prentice-Hall, Upper Saddle River, NJ. ISBN: 978-0131011717, 592 pages.

Knuth, D. E. (1997). *The Art of Computer Programming, Volume 3: Sorting and Searching*. Addison-Wesley, Reading, Massachusetts. ISBN: 0-201-89685-0.

Lehn, K. and Temme, K.-H. (1996). Fuzzy classification of sites suspected of being contaminated. *Ecological Modelling*, 85(1):51–58. DOI: 10.1016/0304-3800(95)00014-3.

Lein, J. K. (1992). Expressing environmental risk using fuzzy variables: a preliminary examination. *Environmental Professional*, 14(3):257–267.

Limbourg, P. and de Rocquigny, E. (2010). Uncertainty analysis using evidence theory – confronting level-1 and level-2 approaches with data availability and computational constraints. *Reliability Engineering & System Safety*, 95(5):550–564. DOI: 10.1016/j.ress.2010.01.005.

Marseguerra, M. and Zio, E. (2002). *Basics of the Monte Carlo Method with Application to System Reliability*. LiLoLe–Verlag. ISBN: 978-3934447066, 141 pages.

Maxwell, R. M., Pelmulder, S. D., Tompson, A. F. B., and Kastenberg, W. E. (1998). On the development of a new methodology for groundwater-driven health risk assessment. *Water Resources Research*, 34(4):833–847. DOI: 10.1029/97WR03605.

Modarres, M. (2006). *Risk Analysis in Engineering: Probabilistic Techniques, Tools and Trends*. CRC Press. ISBN: 978-1574447941, 424 pages.

Moore, R. E. (1979). *Methods and Applications of Interval Analysis*. SIAM, Philadelphia, PA. ISBN: 978-0-89871-161-5.

Möller, B. (2004). Fuzzy randomness – a contribution to imprecise probability. *ZAMM - Journal of Applied Mathematics and Mechanics / Zeitschrift für Angewandte Mathematik und Mechanik*, 84(10-11):754–764. DOI: 10.1002/zamm.200410153.

Möller, B. and Beer, M. (2004). *Fuzzy Randomness: Uncertainty in Civil Engineering and Computational Mechanics*. Springer, Berlin. ISBN: 978-3642073120, 352 pages.

Möller, B. and Beer, M. (2008). Engineering computation under uncertainty — capabilities of non-traditional models. *Computers & Structures*, 86(10):1024–1041. DOI: 10.1016/j.compstruc.2007.05.041.

Möller, B., Beer, M., Graf, W., and Sickert, J.-U. (2006). Time-dependent reliability of textile-strengthened RC structures under consideration of fuzzy randomness. *Computers & Structures*, 84(8–9):585–603. DOI: 10.1016/j.compstruc.2005.10.006.

Möller, B., Graf, W., and Beer, M. (2003). Safety assessment of structures in view of fuzzy randomness. *Computers & Structures*, 81(15):1567–1582. DOI: 10.1016/S0045-7949(03)00147-0.

NASA (2010). Risk-informed decision making handbook (NASA/SP-2010-576). Technical report, NASA. Available at http://standards.nasa.gov/documents/viewdoc/3315763/3315763.

Pasanisi, A., de Rocquigny, E., Bousquet, N., and Parent, E. (2009). Some useful features of the Bayesian setting while dealing with uncertainties in industrial practice. In *Proceedings of the ESREL 2009 Conference*, pages 1795–1802, Prague, Czech Republic. 7-10 September 2009.

Scandurra, E. (1995). *L'ambiente dell'uomo*. Etas Libri, Milan, Italy.

Sentz, K. and Ferson, S. (2002). Combination of evidence in Dempster-Shafer theory. Technical report, Sandia National Laboratories, Albuquerque, New Mexico. Technical Report SAND 2002-0835. Available at http://www.sandia.gov/epistemic/Reports/SAND2002-0835.pdf.

USNRC (2002). An approach for using probabilistic risk assessment in risk-informed decisions on plant-specific changes to the licensing basis (NUREG-1.174). Technical report, US Nuclear Regulatory Commission, Washington, D.C. Available at http://www.nrc.gov/reading-rm/doc-collections/reg-guides/power-reactors/rg/01-174/.

USNRC (2005). Fire PRA methodology for nuclear power facilities (NUREG/CR-6850). Technical re-

port, US Nuclear Regulatory Commission, Washington, D.C. Available at http://www.nrc.gov/reading-rm/doc-collections/nuregs/contract/cr6850/.

USNRC (2009). Guidance on the treatment of uncertainties associated with PRAs in risk-informed decision making (NUREG-1855). Technical report, US Nuclear Regulatory Commission, Washington, D.C. Available at http://www.nrc.gov/reading-rm/doc-collections/nuregs/staff/sr1855/v1/sr1855v1.pdf.

Youngblood, R. W. (2001). Risk significance and safety significance. *Reliability Engineering & System Safety*, 73(2):121–136. DOI: 10.1016/S0951-8320(01)00056-4.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, 8(3):338–353. DOI: 10.1016/S0019-9958(65)90241-X.

Zio, E. (2007). *An introduction to the basics of reliability and risk analysis*. World Scientific. ISBN: 978-9812706393, 180 pages.

Zio, E. and Pedroni, N. (2013). Literature review of methods for representing uncertainty. Cahiers de la Sécurité Industrielle 2013-03, Fondation pour une Culture de Sécurité Industrielle, Toulouse, France. Available at http://www.foncsi.org/.

You can save these bibliographic entries in BibTeX format by clicking on the paper clip icon to the left.

Reproducing this document

You can download this document, and others in the *Cahiers de la Sécurité Industrielle* collection, from FonCSI's web site. Documents are available in PDF, EPUB (for tablets and e-readers) and MOBI (for Kindle e-readers). Paper versions can be ordered online from a print-on-demand service.

Foundation for an Industrial Safety Culture
a public interest research foundation
http://www.FonCSI.org/

6 allée Émile Monso – BP 34038
31029 Toulouse cedex 4
France

Telephone: +33 534 32 32 00
Twitter: @TheFonCSI
Email: contact@FonCSI.org

ISSN 2100-3874

6 allée Émile Monso
ZAC du Palays - BP 34038
31029 Toulouse cedex 4

www.foncsi.org