Propagation of aleatory and epistemic uncertainties in the model for the design of a flood protection dike

Nicola Pedroni, Enrico Zio, Elisa Ferrario, Alberto Pasanisi, Mathieu Couplet

Propagation of aleatory and epistemic uncertainties in the model for the design of a flood protection dike

Nicola Pedroni[a*], Enrico Zio[a,b,c], Elisa Ferrario[b], Alberto Pasanisi[d], Mathieu Couplet[d]
[a]Politecnico di Milano, Milano, Italy
[b]École Centrale Paris, Chatenay Malabry, France
[c]Supélec, Gif-Sur-Yvette, France
[d]Electricité de France, Chatou, France

Abstract: Traditionally, probability distributions are used in risk analysis to represent the uncertainty associated to random (aleatory) phenomena. The parameters (e.g., their mean, variance, …) of these distributions are usually affected by epistemic (state-of-knowledge) uncertainty, due to limited experience and incomplete knowledge about the phenomena that the distributions represent: the uncertainty framework is then characterized by two hierarchical levels of uncertainty. Probability distributions may be used to characterize also the epistemic uncertainty affecting the parameters of the probability distributions. However, when sufficiently informative data are not available, an alternative and proper way to do this might be by means of possibilistic distributions.
In this paper, we use probability distributions to represent aleatory uncertainty and possibility distributions to describe the epistemic uncertainty associated to the poorly known parameters of such probability distributions. A hybrid method is used to hierarchically propagate the two types of uncertainty. The results obtained on a risk model for the design of a flood protection dike are compared with those of a traditional, purely probabilistic, two-dimensional (or double) Monte Carlo approach. To the best of the authors' knowledge, this is the first time that a hybrid Monte Carlo and possibilistic method is tailored to propagate the uncertainties in a risk model when the uncertainty framework is characterized by two hierarchical levels.
The results of the case study show that the hybrid approach produces risk estimates that are more conservative than (or at least comparable to) those obtained by the two-dimensional Monte Carlo method.

Keywords: hierarchical levels of uncertainty; possibility distributions; epistemic dependence

1. INTRODUCTION

In risk analysis, uncertainty is typically distinguished into two types: randomness due to inherent variability in the system behavior (objective, aleatory, stochastic uncertainty) and imprecision due to lack of knowledge and information on the system (subjective, epistemic, state-of-knowledge uncertainty) (Apostolakis, 1990).
We are interested in the framework of two hierarchical levels of uncertainty, referred to as "level-2" setting (Limbourg and de Rocquigny, 2010): the models of the aleatory events (e.g., the failure of a mechanical component, the variation of its geometrical dimensions and material properties, …) contain parameters (e.g., probabilities, failure rates, …) that are epistemically-uncertain, i.e., known with poor precision.

In current risk analysis, both types of uncertainty are represented by means of probability distributions (USNRC, 2009). In such a case, the uncertainty propagation can be carried out by a two-dimensional (or double) Monte Carlo (MC) approach (Rao et al., 2007). However, in some situations, the lack of complete knowledge, information and data impairs the probabilistic representation of epistemic uncertainty. A number of alternative representation frameworks have been proposed to handle such cases (Aven and Zio, 2010), e.g., fuzzy set theory (Klir and Yuan, 1995), evidence theory (Helton et al., 2008), possibility theory (Baudrit et al., 2006 and 2008) and interval analysis (Ferson et al., 2007).

In this paper, we use probability distributions to describe aleatory uncertainty and possibility distributions to describe the epistemic uncertainty in the parameters of the (aleatory) probability distributions. For the propagation of this hybrid (probabilistic and possibilistic) uncertainty representation, the MC technique (Kalos and Withlock, 1986) is combined with the extension principle of fuzzy set theory (Zadeh, 1965) in a "level-2" hierarchical setting (Baudrit et al., 2008). This is done by i) a fuzzy interval analysis to process the uncertainty described by possibility distributions and ii) a repeated MC sampling of the random variables to process aleatory uncertainty (Baudrit et al., 2008; Baraldi and Zio, 2008).

The joint hierarchical propagation of probabilistic and possibilistic representations of uncertainty is applied to a risk model for the design of a flood protection dike (Limbourg and de Rocquigny, 2010); the effectiveness of the hybrid method is compared to that of a traditional two-dimensional MC approach. To the best of the authors' knowledge, this is the first time that a hybrid Monte Carlo and possibilistic method is embraced to propagate the uncertainties in a *risk model* when the uncertainty framework is characterized by two *hierarchical* levels.

The remainder of the paper is organized as follows. In Section 2, the hybrid method for uncertainty propagation is described; in Section 3, the flood model considered for the uncertainty propagation task is presented; in Section 4, the results of the joint hierarchical propagation of aleatory and epistemic uncertainties through the flood model of Section 3 and the comparison with the two-dimensional MC approach are reported and commented; in Section 5, some conclusions are provided.

2. JOINT HIERARCHICAL PROPAGATION OF ALEATORY AND EPISTEMIC UNCERTAINTIES IN A "LEVEL-2" FRAMEWORK

We consider a model whose output is a function $Z = f(Y_1, Y_2, ..., Y_n)$ of n uncertain variables Y_j, j = 1, 2, …, n, whose uncertainty is described by probability distributions $p_{\boldsymbol{\theta}_1}^{Y_1}(y_1), p_{\boldsymbol{\theta}_2}^{Y_2}(y_2), ..., p_{\boldsymbol{\theta}_j}^{Y_j}(y_j), ..., p_{\boldsymbol{\theta}_k}^{Y_k}(y_k)$, where $\boldsymbol{\theta}_j = \{\theta_{j,1}, \theta_{j,2}, ..., \theta_{j,m_j}\}$, j = 1, 2, …, n, are the vectors of the corresponding internal parameters. In a "level-2" framework, the parameters $\boldsymbol{\theta}_j = \{\theta_{j,1}, \theta_{j,2}, ..., \theta_{j,m_j}\}$, j = 1, 2, …, n, of the probability distributions $\{p_{\boldsymbol{\theta}_j}^{Y_j}(y_j):\ j = 1, 2, \cdots, k\}$ are affected by epistemic uncertainty. We describe these uncertainties by possibility distributions $\boldsymbol{\pi}^{\boldsymbol{\theta}_j}(\boldsymbol{\theta}_j) = \left\{\pi^{\theta_{j,1}}(\theta_{j,1}), \pi^{\theta_{j,2}}(\theta_{j,2}), ..., \pi^{\theta_{j,m_j}}(\theta_{j,m_j})\right\}$, j = 1, 2, …, n: the rationale for this choice 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, that are special cases of plausibility and belief functions), which represents the expert's inability to select a *single* probability distribution and, thus, the *imprecision* in his/her knowledge of the uncertain parameters.

In extreme synthesis, the propagation of the hybrid uncertainty information can be performed by combining the Monte Carlo (MC) technique (Kalos and Withlock, 1986) with the extension principle of fuzzy interval analysis (Zadeh, 1965) by means of the following main steps (Baudrit et al., 2008):

1. select one possibility value $\alpha \in (0, 1]$ and the corresponding cuts $A_\alpha^{\theta_{j,1}}, A_\alpha^{\theta_{j,2}}, ..., A_\alpha^{\theta_{j,m_j}}$, j = 1, 2, …, n, of the possibility distributions $\boldsymbol{\pi}^{\theta_j}(\boldsymbol{\theta}_j)$ = $\left\{\pi^{\theta_{j,1}}(\theta_{j,1}), \pi^{\theta_{j,2}}(\theta_{j,2}), ..., \pi^{\theta_{j,m_j}}(\theta_{j,m_j})\right\}$ of the epistemically-uncertain parameters $\boldsymbol{\theta}_j$, j = 1, 2, …, n;
2. randomly sample m intervals $\lfloor \underline{y}_{j,\alpha}^i, \overline{y}_{j,\alpha}^i \rfloor$, i = 1, 2, …, m, j = 1, 2, …, n, of the "probabilistic" variables Y_j, j = 1, 2, …, n, from the probability distributions $\{p_{\boldsymbol{\theta}_j}^{Y_j}(y_j):\ j = 1, 2, ..., n\}$, letting the epistemically-uncertain parameters $\boldsymbol{\theta}_j = \{\theta_{j,1}, \theta_{j,2}, ..., \theta_{j,m_j}\}$ range within the corresponding α-cuts $A_\alpha^{\theta_{j,1}}, A_\alpha^{\theta_{j,2}}, ..., A_\alpha^{\theta_{j,m_j}}$, j = 1, 2, …, n (found at step 1. above);
3. repeat step 2. above for another possibility value $\alpha \in (0, 1]$.

For clarity, in Figure 1 the procedure for sampling the $i-th$ random interval $\lfloor \underline{y}_{j,\alpha}^i, \overline{y}_{j,\alpha}^i \rfloor$ for the generic uncertain variable Y_j is illustrated. Let us suppose that the probability distribution of Y_j is normal with parameters $\boldsymbol{\theta}_j = \{\theta_{j,1}, \theta_{j,2}\} = \{\mu, \sigma\}$; the mean $\mu = \theta_{j,1}$ is represented by a triangular possibility distribution with core c = 5 and support [a, b] = [4, 6] and the standard deviation $\sigma = \theta_{j,2}$ is a fixed point-wise value ($\sigma = \theta_{j,2} = 4$). With reference to the operative procedure outlined above, a possibility value α (e.g., $\alpha = 0.3$ in Figure 1 left) is selected and the corresponding α-cut for $\mu = \theta_{j,1}$ is found, i.e., $\lfloor \underline{\mu}_\alpha, \overline{\mu}_\alpha \rfloor = \lfloor \underline{\theta}_{j,1,\alpha}, \overline{\theta}_{j,1,\alpha} \rfloor$ = [4.3, 5.7] (step 1. above). The cumulative distribution functions $F_{\boldsymbol{\theta}_j}^{Y_j}$ are constructed using the upper and lower values of μ, i.e., $\underline{\mu}_\alpha = \underline{\theta}_{j,1,\alpha} = 4.3$ and $\overline{\mu}_\alpha = \overline{\theta}_{j,1,\alpha} = 5.7$ (Figure 1 right); then, a random number u_j^i (e.g., $u_j^i = 0.7$ in Figure 1 right) is sampled from a uniform distribution in [0,1) and the interval $\lfloor \underline{y}_{j,\alpha}^i, \overline{y}_{j,\alpha}^i \rfloor$ is computed as $\left[\inf_{\theta_j \in [\underline{\theta}_{j,\alpha}, \overline{\theta}_{j,\alpha}]} \left[F_{\theta_j}^{Y_j}\right]^{-1}\left(u_j^i\right), \sup_{\theta_j \in [\underline{\theta}_{j,\alpha}, \overline{\theta}_{j,\alpha}]} \left[F_{\theta_j}^{Y_j}\right]^{-1}\left(u_j^i\right)\right] = \left[\inf_{\mu \in [\underline{\mu}_\alpha, \overline{\mu}_\alpha]} \left[F_{\mu}^{Y_j}\right]^{-1}\left(u_j^i\right), \sup_{\mu \in [\underline{\mu}_\alpha, \overline{\mu}_\alpha]} \left[F_{\mu}^{Y_j}\right]^{-1}\left(u_j^i\right)\right] =$ $\left[\inf_{\mu \in [4.3, 5.7]} \left[F_{\mu}^{Y_j}\right]^{-1}(0.7), \sup_{\mu \in [4.3, 5.7]} \left[F_{\mu}^{Y_j}\right]^{-1}(0.7)\right] = [6.4, 7.8]$ (step 2. above).

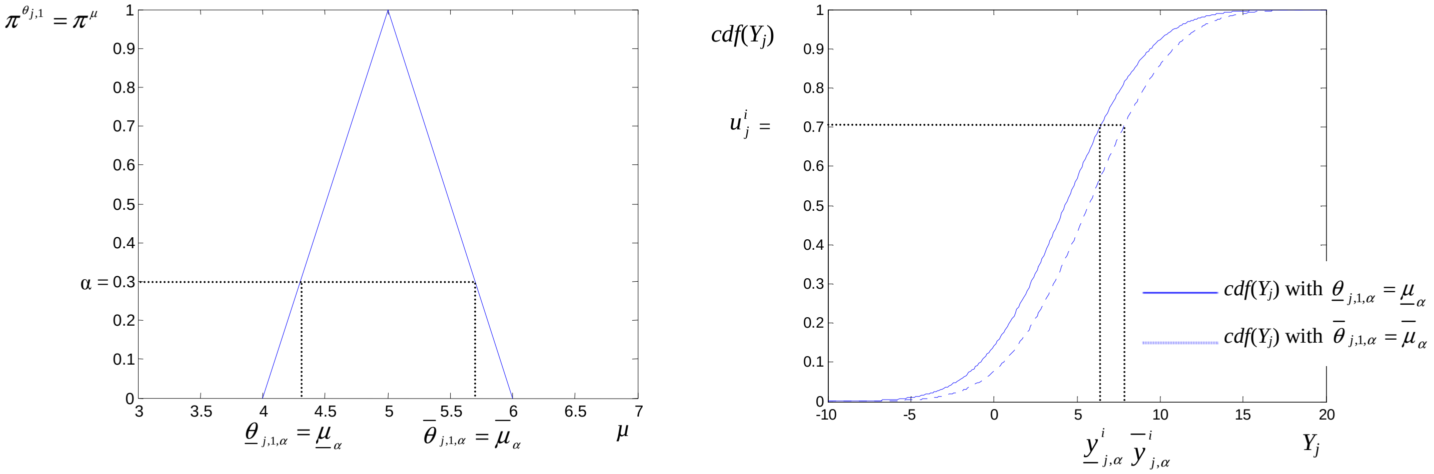

Figure 1. Left: triangular possibility distribution of the mean μ of the normal probability distribution of $Y_j \sim N(\mu, 4) = N(\boldsymbol{\theta})$; in evidence the α-cut of level $\alpha = 0.3$ $\left[\underline{\theta}_{j,1,\alpha}, \overline{\theta}_{j,1,\alpha}\right] = \left[\underline{\mu}_\alpha, \overline{\mu}_\alpha\right] = [4.3, 5.7]$. Right: cumulative distribution functions of Y_j built in correspondence of the extreme values $\underline{\mu}_\alpha = 4.3$ and $\overline{\mu}_\alpha = 5.7$ of the α-cut $\left[\underline{\mu}_\alpha, \overline{\mu}_\alpha\right]$ of μ.

For each set A contained in the universe of discourse U_Z of the output variable Z, the output of the algorithm is represented by a set of plausibility functions $\{Pl_\alpha(A) : \alpha \in (0,1]\}$ and a set of belief functions $\{Bel_\alpha(A) : \alpha \in (0,1]\}$, obtained in correspondence of the *different* possibility values $\alpha \in (0, 1]$ selected at step 1. above; these sets of functions are then synthesized into the plausibility $Pl(A)$ and belief $Bel(A)$ of A as $\int_0^1 Pl_\alpha(A)d\alpha$ and $\int_0^1 Bel_\alpha(A)d\alpha$, respectively (Baudrit et al., 2006). Notice that $Pl(A)$ and $Bel(A)$ bound above and below, respectively, the probability $P(A)$ of A, i.e., $Bel(A) \leq P(A) \leq Pl(A)$. 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, $Bel(f(Y) \in (-\infty, z])$ and $Pl(f(Y) \in (-\infty, z])$ can be interpreted as bounding cumulative distributions $\underline{F}(z) = Bel(f(Y) \in (-\infty, z])$, $\overline{F}(z) = Pl(f(Y) \in (-\infty, z])$.

Finally, it is worth noting that performing an interval analysis on α-cuts assumes *total dependence* between the uncertain parameters. Actually, this procedure implies strong dependence between the information sources (e.g., the experts or observers) that supply the input possibility distributions, because the same *confidence level* ($1 - \alpha$) is chosen to build the α-cuts for all the uncertain parameters (Baudrit et al., 2006).

3. CASE STUDY: FLOOD PROTECTION DESIGN

3.1. The model

The model considered calculates the maximal water level of the river (i.e., the output variable of the model, Z_c), given several parameters (i.e., the input variables of the model) (Limbourg and de Rocquigny, 2010):

$$Z_c = Z_v + \left(\frac{Q}{K_s * B * \sqrt{(Z_m - Z_v)/L}} \right)^{3/5} \tag{1}$$

where: Q is the yearly maximal water discharge (m^3/s); Z_m and 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_v, K_s.

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

In this Section, a detailed description of the uncertain input variables is given together with the explanation of the reasons underlying the choices of their description by probability and possibility distributions.

3.2.1. The yearly maximal water flow, *Q*

The Gumbel distribution $Gum(q|\alpha,\beta)$ is a well-established probabilistic (aleatory) model for maximal flows (Limbourg and de Rocquigny, 2010):

$$Gum(q|\alpha,\beta)=\frac{1}{\beta}\exp\left[-\exp\left(\frac{q-\alpha}{\beta}\right)\right]\exp\left[\frac{\alpha-q}{\beta}\right] \tag{2}$$

The extreme physical bounds on variable Q are (Limbourg and de Rocquigny, 2010): $Q_{\min}=10$ m^3/s, which is a typical drought flow level (irrelevant within a flood study); $Q_{\max}=10000$ m^3/s, which is three times larger than the maximal flood ever occurred.

The parameters α and β in (2) are affected by epistemic uncertainty; however, a large amount of data (i.e., 149 annual maximal flow values) is available for performing statistical inference on them. In particular, the point estimates $\hat{\mu}_\alpha$ and $\hat{\mu}_\beta$ and the corresponding standard deviations $\hat{\sigma}_\alpha$ and $\hat{\sigma}_\beta$ have been obtained for the parameters α and β of the Gumbel distribution (2) by performing Maximum Likelihood Estimations (MLEs) with the 149 data available: the method has provided $\hat{\mu}_\alpha=1013$ m^3/s, $\hat{\mu}_\beta=558$ m^3/s, $\hat{\sigma}_\alpha=48$ m^3/s and $\hat{\sigma}_\beta=36$ m^3/s (Limbourg and de Rocquigny, 2010). A probabilistic treatment of this epistemic uncertainty has been proposed in the original paper by Limbourg and de Rocquigny (2010): in particular, $\alpha\sim p^\alpha(\alpha)=N(\hat{\mu}_\alpha,\hat{\sigma}_\alpha)=N(1013,48)$ and $\beta\sim p^\beta(\beta)=N(\hat{\mu}_\beta,\hat{\sigma}_\beta)=N(558,36)$.

In this paper, the Gumbel shape of the aleatory probability distributions (2) is retained but the epistemic uncertainty on the parameters is represented in possibilistic terms. To do so the normal probability distributions $p^\alpha(\alpha)$ and $p^\beta(\beta)$ used in Limbourg and de Rocquigny (2010) are transformed into the possibility distributions $\pi^\alpha(\alpha)$ and $\pi^\beta(\beta)$ by *normalization*, i.e., $\pi^\alpha(\alpha)=\frac{p^\alpha(\alpha)}{\sup p^\alpha(\alpha)}$, $\pi^\beta(\beta)=\frac{p^\beta(\beta)}{\sup p^\beta(\beta)}$. The supports of the possibility distributions $\pi^\alpha(\alpha)$ and $\pi^\beta(\beta)$ are set to $[\hat{\mu}_\alpha-\hat{\sigma}_\alpha,\hat{\mu}_\alpha+\hat{\sigma}_\alpha]=[965,1061]$ and $[\hat{\mu}_\beta-\hat{\sigma}_\beta,\hat{\mu}_\beta+\hat{\sigma}_\beta]=[523,594]$, respectively, according to the suggestions by Limbourg and de Rocquigny (2010).

3.2.2. The upstream and downstream riverbed levels, Z_m and Z_v

The minimum and maximum physical bounds on variables Z_m and Z_v are $Z_{m,\min}=53.5$ m and $Z_{v,\min}=48$ m (given by plausible lower geomorphologic limits to erosion) and $Z_{m,\max}=57$ m and $Z_{v,\max}=51$ m (given by plausible upper geomorphologic limits to sedimentation), respectively. Normal distributions truncated at the minimum and maximum physical bounds have been selected in Limbourg and de Rocquigny (2010) to represent the aleatory part of the uncertainty, i.e., $Z_m\sim N(\mu_{Zm},\sigma_{Zm})$ and $Z_v\sim N(\mu_{Zv},\sigma_{Zv})$. An amount of 29 data has been used in the reference paper by Limbourg and de Rocquigny (2010) to provide the point estimates $\hat{\mu}_{Zm}=55.03$ m, $\hat{\mu}_{Zv}=50.19$ m, $\hat{\sigma}_{Zm}=0.45$ m, $\hat{\sigma}_{Zv}=0.38$ m for parameters μ_{Zm}, μ_{Zv}, σ_{Zm} and σ_{Zv}, respectively, by means of the MLE method. However, according to Limbourg and de Rocquigny (2010) there is large uncertainty about the shape of the probability distributions of Z_m and Z_v; as a consequence the authors embrace a conservative "level-2", using the MLE method to provide also standard deviations as a measure of the uncertainty on the point estimates $\hat{\mu}_{Zm}$, $\hat{\mu}_{Zv}$, $\hat{\sigma}_{Zm}$ and $\hat{\sigma}_{Zv}$: in particular, $\hat{\sigma}_{\hat{\mu}_{Zm}}=0.08$, $\hat{\sigma}_{\hat{\mu}_{Zv}}=0.07$, $\hat{\sigma}_{\hat{\sigma}_{Zm}}=0.06$ and $\hat{\sigma}_{\hat{\sigma}_{Zv}}=0.05$. Using this information, Limbourg and de Rocquigny (2010) model the epistemic uncertainty associated to the parameters μ_{Zm}, μ_{Zv}, σ_{Zm} and σ_{Zv} by normal distributions, i.e., $\mu_{Zm}\sim N(\hat{\mu}_{Zm},\hat{\sigma}_{\hat{\mu}_{Zm}})$, $\mu_{Zv}\sim N(\hat{\mu}_{Zv},\hat{\sigma}_{\hat{\mu}_{Zv}})$, $\sigma_{Zm}\sim N(\hat{\sigma}_{Zm},\hat{\sigma}_{\hat{\mu}_{Zm}})$ and $\sigma_{Zv}\sim N(\hat{\sigma}_{Zv},\hat{\sigma}_{\hat{\sigma}_{Zv}})$.

In this paper, the shapes of the aleatory probability distributions for Z_m and Z_v, i.e., $N(\mu_{Zm},\sigma_{Zm})$ and $N(\mu_{Zv},\sigma_{Zv})$, are kept unaltered with respect to those of Limbourg and de Rocquigny (2010); on the contrary, the information produced by the MLE method on parameters μ_{Zm}, μ_{Zv}, σ_{Zm} and σ_{Zv}, i.e., the point estimates

$\hat{\mu}_{Zm}$, $\hat{\mu}_{Zv}$, $\hat{\sigma}_{Zm}$, $\hat{\sigma}_{Zv}$ and the corresponding standard deviations $\hat{\sigma}_{\hat{\mu}_{Zm}}$, $\hat{\sigma}_{\hat{\mu}_{Zv}}$, $\hat{\sigma}_{\hat{\sigma}_{Zm}}$, $\hat{\sigma}_{\hat{\sigma}_{Zv}}$, is used to build possibility distributions for μ_{Zm}, μ_{Zv}, σ_{Zm} and σ_{Zv} by means of the Chebyshev inequality (Baudrit and Dubois, 2006). The classical Chebyshev inequality 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:

$$P(|Y-\mu| \leq k\sigma) \geq 1-\frac{1}{k^2} \text{ for } k \geq 1. \quad (3)$$

Formula (3) can be thus used to define a possibility distribution π that dominates any probability density function with given mean μ and standard deviation σ by considering intervals $[\mu-k\sigma, \mu+k\sigma]$ as *α*-cuts of π and letting $\pi(\mu-k\sigma)=\pi(\mu+k\sigma)=\frac{1}{k^2}=\alpha$. This possibility distribution defines a probability family which has been proven to contain *all* probability distributions with mean μ and standard deviation σ (Baudrit and Dubois, 2006).

In this case, the point estimates $\hat{\mu}_{Zm}$, $\hat{\mu}_{Zv}$, $\hat{\sigma}_{Zm}$ and $\hat{\sigma}_{Zv}$ produced by the MLE method, are used in (3) as the means of the parameters μ_{Zm}, μ_{Zv}, σ_{Zm} and σ_{Zv}, whereas the errors $\hat{\sigma}_{\hat{\mu}_{Zm}}$, $\hat{\sigma}_{\hat{\mu}_{Zv}}$, $\hat{\sigma}_{\hat{\sigma}_{Zm}}$ and $\hat{\sigma}_{\hat{\sigma}_{Zv}}$ associated to the estimates $\hat{\mu}_{Zm}$, $\hat{\mu}_{Zv}$, $\hat{\sigma}_{Zm}$ and $\hat{\sigma}_{Zv}$ are used in (3) as the standard deviations of the parameters μ_{Zm}, μ_{Zv}, σ_{Zm} and σ_{Zv} in order to build the corresponding possibility distributions $\pi^{\mu_{Zm}}$, $\pi^{\mu_{Zv}}$, $\pi^{\sigma_{Zm}}$ and $\pi^{\sigma_{Zv}}$; the supports of the possibility distributions are obtained by extending two times the standard deviation $\hat{\sigma}_{\hat{\mu}_{Zm}}$, $\hat{\sigma}_{\hat{\mu}_{Zv}}$, $\hat{\sigma}_{\hat{\sigma}_{Zm}}$ and $\hat{\sigma}_{\hat{\sigma}_{Zv}}$ in both directions with respect to the estimates $\hat{\mu}_{Zm}$, $\hat{\mu}_{Zv}$, $\hat{\sigma}_{Zm}$ and $\hat{\sigma}_{Zv}$.

3.2.3. The Strickler friction coefficient, K_S

The Strickler friction coefficient K_s is perhaps the most critical source of uncertainty: actually, it is a simplification of a much more complex hydraulic model. The absolute physical limits of K_s are $[a,b]=[5,60]$ (Limbourg and de Rocquigny, 2010): $K_s < 5$ corresponds to an extremely sinuous shape of the canal, with large dents and strong vegetation; $K_s = 60$ corresponds to a canal with smoothest earth surface, rectilinear, without any vegetation.

There is an underlying natural variability in the friction coefficient K_s since it is affected by unpredictable events modifying the river status (erosion/sedimentation, etc.): this variability is plausibly inferred as a normal distribution, i.e., $K_s \sim N(\mu_{Ks}, \sigma_{Ks})$ (Limbourg and de Rocquigny, 2010). Unfortunately, the mean value μ_{Ks} of this Gaussian distribution is highly uncertain and difficult to measure; actually, direct measurement is impossible and data may only be retrieved through indirect calibration noised by significant observational uncertainty: this is reflected in only a very small series of 5 data sets available with ± 15% noise (Limbourg and de Rocquigny, 2010). The sample mean $\hat{\mu}_{Ks}$ and standard deviation $\hat{\sigma}_{Ks}$ of these five data sets equal 27.8 and 3, respectively. In order to reflect the full amount of imprecision generated by the indirect measurement, the minimal sample mean $\hat{\mu}_{\min} = 23.63$ and the maximal sample mean $\hat{\mu}_{\max} = 31.97$ are also calculated, all measurements being conservatively assumed to be biased in the same direction. Moreover, since the low sample size adds another sort of "statistical" epistemic uncertainty to the values $\hat{\mu}_{\min}$ and $\hat{\mu}_{\max}$, the 70% confidence bounds on the mean estimates $\hat{\mu}_{\min}$ and $\hat{\mu}_{\max}$ are also computed as $\hat{\mu}_{\min} - \frac{\hat{\sigma}_{Ks}}{\sqrt{5}} = 22.3$ and $\hat{\mu}_{\min} - \frac{\hat{\sigma}_{Ks}}{\sqrt{5}} = 33.3$, respectively. In Limbourg and de Rocquigny (2010), these considerations result in the following uncertainty quantification for K_s:

$$K_s \sim N(\mu_{Ks}, \sigma_{Ks}), \text{ with } \sigma_{Ks} = \hat{\sigma}_{Ks} = 3 \text{ and } \mu_{Ks} \in \left[\hat{\mu}_{\min} - \frac{\hat{\sigma}_{Ks}}{\sqrt{5}}, \hat{\mu}_{\max} + \frac{\hat{\sigma}_{Ks}}{\sqrt{5}}\right] = [22.3, 33.3]. \quad (4)$$

In this paper, the shape of the aleatory probability distribution of K_s, i.e., $N(\mu_{Ks}, \sigma_{Ks})$ in (4) is retained; however, differently from the original paper, a possibility distribution is associated to μ_{Ks}. In particular, a *trapezoidal* possibility distribution is here proposed: the support is chosen to be

$[a,b]=\left[\hat{\mu}_{\min}-\frac{\hat{\sigma}_{Ks}}{\sqrt{5}},\hat{\mu}_{\max}+\frac{\hat{\sigma}_{Ks}}{\sqrt{5}}\right]=[22.3,33.3]$ as in (4); however, in this paper additional information is provided concerning the most likely values of μ_{Ks} exploiting the available data set: in particular, since the core of the trapezoidal distribution contains the most likely values of the parameter μ_{Ks}, in this case it is set to $[c,d]=\left[\hat{\mu}_{K_s}-\frac{\hat{\sigma}_{Ks}}{\sqrt{5}},\hat{\mu}_{K_s}+\frac{\hat{\sigma}_{Ks}}{\sqrt{5}}\right]=[26.5,29.1]$, i.e., the interval obtained by adding/subtracting to the sample mean $\hat{\mu}_{Ks}=27.8$ (which is assumed to be the *most likely value* for μ_{Ks}) the "statistical" epistemic uncertainty due to the low sample size (i.e., the quantity $\frac{\hat{\sigma}_{Ks}}{\sqrt{5}}$).

4. APPLICATION

In this Section, the following approaches are considered and compared in the task of hierarchically propagating aleatory and epistemic uncertainties in a "level-2" framework: i) the hybrid Monte Carlo (MC) and possibilistic approach of Section 2; ii) a two-dimensional (double) MC approach: a) assuming *independence* between the epistemically-uncertain parameters of the aleatory probability distributions; b) assuming *total dependence* between the epistemically-uncertain parameters of the aleatory probability distributions. This choice has been made to perform a fair comparison with the hybrid MC and possibilistic approach, which implicitly assumes by construction total dependence between the epistemically-uncertain parameters (see Section 2)[1].

It is worth noting that the probability distributions here used in the two-dimensional MC approach for Q, Z_m and Z_v and for the corresponding epistemically-uncertain parameters are the same as those proposed in the original paper by Limbourg and de Rocquigny (2010) (and recalled in Section 3.2.1 and 3.2.2); the only exception is represented by the probability distribution for μ_{Ks}, which for consistency and coherence of the comparison is here obtained by normalization of the trapezoidal possibility distribution described in Section 3.2.3, i.e., $p^{\mu_{Ks}}(\mu_{Ks})=\frac{\pi^{\mu_{Ks}}(\mu_{Ks})}{\int_a^b \pi^{\mu_{Ks}}(\mu_{Ks})\,d\mu_{Ks}}$.

For simplicity, we start by comparing approaches ii.a and ii.b. above., i.e., double MC assuming independence and total dependence between the uncertain parameters, respectively. Figure 2 top left shows the upper and lower cumulative distribution functions of the model output Z_c obtained by the double MC approach assuming independence (ii.a) and total dependence (ii.b), respectively. In this case, assuming total dependence between the uncertain parameters is shown to lead to a smaller gap between the upper and lower cumulative distribution functions of the model output Z_c than assuming independence. This can be easily explained by analyzing the input-output functional relationship considered, i.e., $Z_c=Z_v+\left(\frac{Q}{K_s * B * \sqrt{(Z_m-Z_v)/L}}\right)^{3/5}$ in (1): it can be seen that one of the input variables (i.e., Q) appears at the numerator of the expression, whereas others (i.e., K_s and Z_m) appear at the denominator, and another one appears both at the numerator and at the denominator (i.e., Z_v). In such a case, the highest possible values for the model output Z_c are obtained with a *combination* of high values of *both* Q *and* Z_v (i.e., in other words, high values of the corresponding uncertain parameters α, β, μ_{Zv} and σ_{Zv}) *and* low values of *both* K_s *and* Z_m (i.e., in other words, low values of the corresponding uncertain parameters μ_{Ks}, σ_{Ks}, μ_{Zm} and σ_{Zm}); conversely, the lowest possible values for the model output Z_c are obtained with a combination of low values of both Q and Z_v and high values of both K_s and Z_m. These extreme situations (which give rise to

---

[1] It is very important to note that the condition of total epistemic (or state-of-knowledge) dependence between parameters of risk models is far from unlikely. For example, consider the case of a system containing a number of physically distinct, but similar/ nominally identical components whose failure rates are estimated by means of the same data set: in such situation, the distributions describing the uncertainty associated to the failure rates have to be considered totally dependent (Apostolakis and Kaplan, 1981; USNRC, 2009).

the largest separation between the upper and lower cumulative distribution functions, i.e., to the most "epistemically-uncertain" and, thus, conservative case), can be obtained only in case ii.a above, i.e., assuming independence between the epistemically-uncertain parameters. Actually, if a pure random sampling is performed among independent uncertain parameters, *all* possible *combinations* of values can be in principle generated, since the entire ranges of variability of the uncertain parameters can be explored independently: thus, in some random samples, high values of Q and Z_v may be combined by chance with low values of both K_s and Z_m, whereas in other random samples low values of both Q and Z_v may be combined by chance with high values of both K_s and Z_m. Conversely, such "extreme" situations cannot occur if there is total dependence between the uncertain parameters (i.e., case ii.b above). Actually, in such a case high (low) values of both Q and Z_v can *only* be combined with high (low) values of both K_s and Z_m, giving rise to values of output Z_c which are lower (higher) than the highest (lowest) possible: in other words, the separation between the upper and lower cumulative distribution functions produced in case ii.b is *always smaller* than that produced by the "extreme" situations described above (which are possible *only* in case ii.a).

A final, straightforward remark is in order. The considerations made above about what combinations of parameter values would lead to the most conservative results (i.e., to the largest gap between the upper and lower cumulative distribution functions) are strictly dependent on the input-output relationship considered: obviously, a *different model* (with different functional relationships between inputs and outputs) would require *different combinations* of input values in order to obtain the most conservative results. For example, for the hypothetical model $w = (x * y)/z$ the most conservative results (i.e., the largest separation between the upper and lower cumulative distribution functions) would be obtained by imposing *total* dependence between x and y and *opposite* dependence between z and both x and y.

We now move on to compare i. and ii.a. Figure 2 top right shows the upper and lower cumulative distribution functions of the model output Z_c obtained by the double MC approach assuming independence between the uncertain parameters (case ii.a) and the plausibility and belief functions produced by the hybrid MC and possibilistic approach (case i). The results are very similar, which is explained as follows. First of all, there is obviously a strong similarity between the shapes of the probability distributions of the epistemically-uncertain parameters used in the double MC approach and the corresponding possibility distributions used in the hybrid approach. For example, the ranges of variability of the uncertain parameters are the same for both the probability and the possibility distributions considered (see Sections 3.2.1-3.2.3); in addition, some of the possibility distributions employed in the hybrid approach (e.g., those of parameters α and β of the Gumbel distribution for Q) are obtained by simple normalization of the probability distributions employed in the double MC approach (Section 3.2.1); finally, the trapezoidal probability distribution used in the double MC approach for the Strickler friction coefficient K_s is obtained by simple normalization of the trapezoidal possibility distribution proposed in the present paper and described in Section 3.2.3. In addition to the similarity between the probability and possibility distributions considered, the second motivation for the similarity between the results lies in the characteristics of the two algorithms used to propagate the uncertainties. In the double MC approach, a plain random sampling is performed from the probability distribution of the epistemically-uncertain parameters, which are considered independent: as a consequence of this independence, in principle *all* possible *combinations* of values of the parameters can be sampled, since the entire ranges of variability of the parameters are explored *randomly* and *independently*. In the hybrid approach, the same confidence level α is chosen to build the α-cuts for all the possibility distributions of the uncertain parameters; then, the minimum and maximum values of the model output Z_c are identified letting the uncertain parameters range *independently* within the corresponding α-cuts (step 2. of the procedure in Section 2): thus, even in this way, once a possibility level α is selected, *all* possible *combinations* of parameter values can be explored, since the α-cuts of all the parameters are *exhaustively* searched to maximize/minimize the model output Z_c.

As final comparison, Figure 2 bottom shows the upper and lower cumulative distribution functions of the model output Z_c obtained by the double MC approach assuming total dependence between parameters (case ii.b) and the hybrid MC approach (case i.) (which assumes total dependence between parameters). From the consideration made above it is clear why the gap is smaller between the cumulative distributions in the two-dimensional MC approach assuming total dependence between the uncertain parameters (case ii.b) than between the plausibility and belief functions produced by the hybrid approach (case i.)[2]: actually, in case ii.b only a limited set of combinations of uncertain parameter values can be randomly explored, whereas in case i. once a value for α is selected, all possible combinations of uncertain parameter values are exhaustively searched to maximize/minimize the output (giving rise to a larger separation between the plausibility and belief functions). In this respect, it is with recalling that the hybrid and the double MC approaches represent the epistemic uncertainty in radically different way: in particular, in the hybrid method, possibility distributions are employed which identify a *family* of probability distributions for the epistemically-uncertain parameters; on the contrary, in the double MC approach, only a *single* probability distribution is assigned to represent the epistemic uncertainty associated to the parameters.

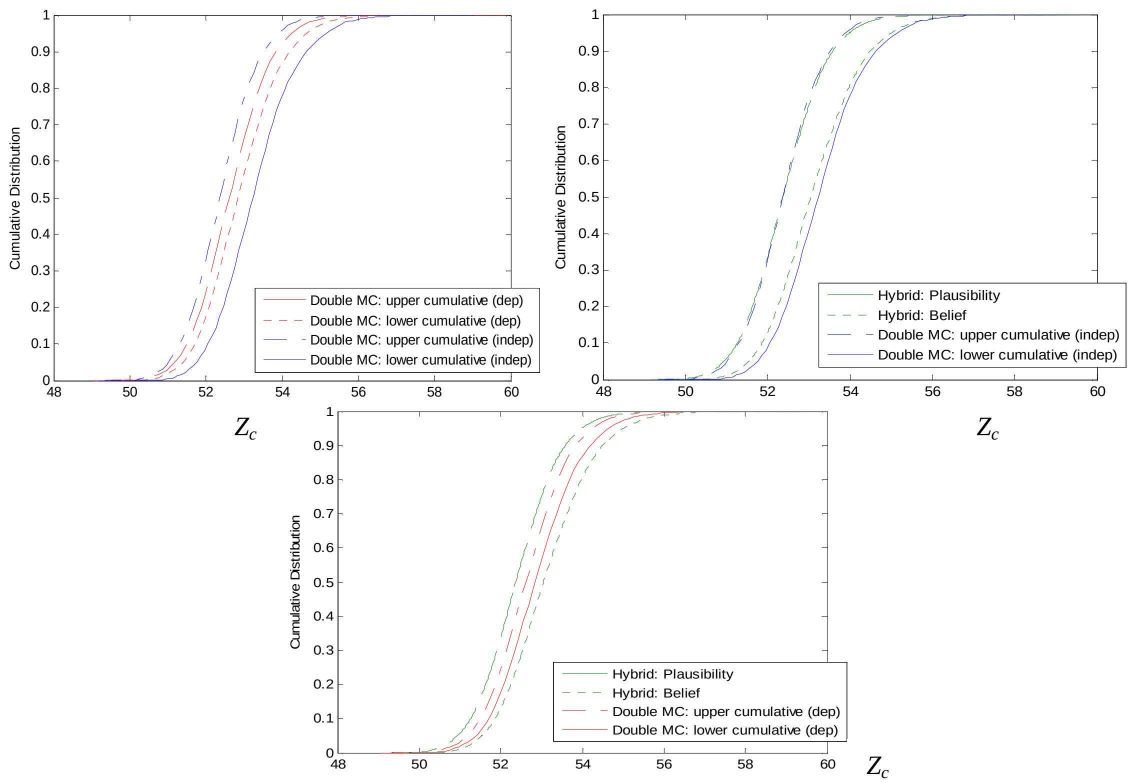

Figure 2. Comparison of the CDFs of Z_c obtained by: i) the two-dimensional MC approach, considering both independence and total dependence between the epistemically-uncertain parameters (top, left); ii) the hybrid MC and possibilistic approach and the two-dimensional MC approach assuming independence between the epistemically-uncertain parameters (top, right); iii) the hybrid method and the two-dimensional MC method assuming total dependence between the epistemically-uncertain parameters (bottom).

A final remark is in order with respect to the results obtained. Since in this case the hybrid MC and possibilistic approach gives rise to a larger separation between the plausibility and belief functions than the double MC approach (assuming total dependence between the epistemically-uncertain parameters), it can be considered *more conservative*. As a consequence, embracing one method instead of the other may significantly change the outcome of a decision making process in a risk assessment problem involving uncertainties: this is of paramount importance in systems that are critical from the safety view point, e.g., in the nuclear, aerospace, chemical and environmental fields. A quantitative demonstration of this statement is given in what follows.

The final goal of the uncertainty propagation 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 most 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 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 most relevant to

[2] As before, notice that this comparison is fair because both methods assume total dependence between the epistemically-uncertain parameters.

the decision maker is the probability that the maximal water level of the river Z_c exceeds a given threshold $z*$, i.e., $P(Z_c \geq z*)$; in the present report, $z*$ = 55.5 m as in Limbourg and de Rocquigny (2010). Table 1 reports the lower ($Z_{c,lower}^{0.99}$) and upper ($Z_{c,upper}^{0.99}$) 99th percentiles obtained from the two limiting cumulative distributions and the corresponding $LowerBound(Z_c \geq z*)$ and $UpperBound(Z_c \geq z*)$. In addition, as synthetic mathematical indicators of the imprecision in the knowledge of Z_c (i.e., of the separation between the lower and upper cumulative distribution functions), the following percentage widths have been reported:

- $W_{Zc} = \frac{Z_{c,upper}^{0.99} - Z_{c,lower}^{0.99}}{Z_{c,prob}^{0.99}}$ of the interval $\left[Z_{c,lower}^{0.99}, Z_{c,upper}^{0.99}\right]$ with respect to the percentile $Z_{c,prob}^{0.99}$ obtained by a traditional, one-dimensional pure probabilistic approach of reference, i.e., an approach where the parameters of the aleatory probability distributions are *fixed*, *known values* (see Limbourg and de Rocquigny (2010) for details);
- $W* = \frac{UpperBound(Z_c \geq z*) - LowerBound(Z_c \geq z*)}{P(Z_c \geq z*)_{prob}}$ of the interval [$LowerBound(Z_c \geq z*)$, $UpperBound(Z_c \geq z*)$] with respect to the probability $P(Z_c > z*)_{prob}$ obtained by a one-dimensional pure probabilistic approach of reference (see Limbourg and de Rocquigny (2010) for details).

Table 1. Comparison of the lower and upper values of Z_c percentiles and threshold exceedance probability obtained by the three method analyzed; the respective percentage widths *W* of the intervals are also reported.

Method	**$Z_c^{0.99}$ [m] (Pure prob. = 55.34m)**		**$P[Z_c \geq 55.5]$(Pure prob. = 0.0076)**	
	$\left[Z_{c,lower}^{0.99}, Z_{c,upper}^{0.99}\right]$	$W_{Zc}[\%]$	**[LowerBound, UpperBound]**	$W*[\%]$
Hybrid MC and possibilistic (total dependence)	[54.79, 56.03]	2.2	[0.0024, 0.0241]	286
Double MC (independence)	[54.56, 56.06]	2.7	[0.0013, 0.0293]	368
Double MC (total dependence)	[54.05, 55.50]	0.8	[0.0042, 0.0111]	91

The considerations reported above are confirmed: there is a similarity between the values of the indicators relative to the hybrid MC and possibilistic approach (case i.), and to the double MC approach assuming independence among the uncertain parameters (case ii.a); on the contrary, there is a significant difference between the values of the indicators relative to the hybrid method and to the double MC approach assuming total dependence between the uncertain parameters (case ii.b). In particular, one additional consideration concerning this latter comparison is worth to be done. Analyzing, for instance, the probability that the maximal water level of the river Z_c exceeds the threshold $z*$ = 55.5 m, $P[Z_c \geq z* = 55.5]$, it can be seen that the hybrid approach is much more conservative than the double MC approach assuming total dependence between parameters: in fact, for instance, the upper bounds of $P[Z_c \geq z*]$ are 0.0241 and 0.0111 for cases i. and ii.b, respectively. Thus, in this case the use of the double MC approach would lead to underestimate by about 54% the probability that the maximal water level of the river Z_c exceeds the threshold $z*$ = 55.5 m: in other words, it would lead to underestimate by about 54% the "failure probability" of the dike and, at the same time, the flood risk. The same consideration holds for the dike level necessary to guarantee a 100 year-return period represented by the 99% quantile $Z_c^{0.99}$ of the water level of the river; for example, the upper bounds of $Z_c^{0.99}$ are 56.03m and 55.50m for cases i. and ii.b, respectively. Thus, also in this case the use of the double MC approach would lead to a slight underestimation of the dike level necessary to guarantee a 100 year flood return period. Therefore, even if the double MC approach purposedly tries to separate variability from imprecision, differently from the hybrid approach, it treats lack-of-knowledge in the same way as it treats variability (i.e., using probability distributions): as a consequence, in some cases, it may fail to produce reliable and conservative results, which can raise great concerns from the safety point of view.

5. CONCLUSIONS

A hybrid method has been applied for the joint propagation of probabilistic and possibilistic uncertainty representations onto a flood model in a "level-2" framework. The results obtained have been compared with those produced by a double MC approach. In particular, the following analyses have been performed:

1. a comparison between two-dimensional MC approaches assuming total dependence and independence between the parameters, respectively, highlighting that in this case study, assuming

independence among parameters leads to a larger gap between the cumulative distributions of the model output than assuming total dependence;

2. a comparison between the hybrid approach and the two-dimensional MC approach assuming independence between the epistemically-uncertain parameters, showing that in this case study, the cumulative distribution functions of the model output produced by the two approaches are similar;
3. a comparison between the hybrid and the two-dimensional MC approach assuming total dependence between the parameters, showing that the gap between the plausibility and belief functions of the model output produced by the hybrid approach is larger than the gap between the upper and lower CDFs produced by the two-dimensional MC method (i.e., the results produced by the hybrid approach are *more conservative*). This has been quantitatively confirmed by way of the risk model for the design of a flood protection dike through the computation of i) the dike level necessary to guarantee a 100 year flood return period and ii) the flood risk for a given dike level. In fact, both quantities have been underestimated by the double MC approach with respect to the hybrid approach.

The considerations above confirm that embracing different methods for jointly propagating aleatory and epistemic uncertainties may generate different results, thus producing significant changes in the outcomes of decision making processes in risk assessment problems involving uncertainties: this is of paramount importance in systems that are critical from the safety view point, e.g., in the nuclear, aerospace, chemical and environmental fields.

It seems advisable to conclude that, if nothing is known about the dependence or independence relationship between the epistemically-uncertain parameters, it may be advisable to resort to the hybrid MC and possibilistic approach because its risk estimates are more conservative than (or at least comparable to) those obtained by the double MC approach assuming dependence (or independence) between the epistemically-uncertain parameters.

6. REFERENCES

Apostolakis, G.E (1990) The Concept of Probability in Safety Assessments of Technological Systems. *Science*, 250, 1359-1364.

Apostolakis, G.E. and Kaplan, S. (1981) Pitfalls in risk calculations. *Reliability Engineering*, 2, 135-145.

Aven, T. and Zio, E. (2010) Some considerations on the treatment of uncertainties in risk assessment for practical decision making. *Reliability Engineering and System Safety*, 96, 64-74.

Baraldi, P. and Zio, E. (2008) A Combined Monte Carlo and Possibilistic Approach to Uncertainty Propagation in Event Tree Analysis. *Risk Analysis*, 28, 1309-1326.

Baudrit, C. and Dubois, D. (2006) Practical Representations of Incomplete Probabilistic Knowledge. *Computational Statistics & Data Analysis*, 51, 86-108.

Baudrit, C., Dubois, D., Guyonnet, D. (2006) Joint Propagation and Exploitation of Probabilistic and Possibilistic Information in Risk Assessment. *IEEE Transactions on Fuzzy Systems*, 14, 593-608.

Baudrit, C., Dubois, D., Perrot, N. (2008) Representing parametric probabilistic models tainted with imprecision. *Fuzzy Sets and System*, 159, 1913-1928.

Ferson, S., Kreinovich, V., Hajagos, J., Oberkampf W., Ginzburg, L. (2007) *Experimental Uncertainty Estimation and Statistics for Data Having Interval Uncertainty*. Setauket, New York, SAND2007-0939.

Helton, J.C., Johnson, J.D., Oberkampf, W.L., Sallaberry, C.J. (2008) *Representation of Analysis Results Involving Aleatory and Epistemic Uncertainty*, SAND2008-4379.

Kalos, M. H. and Whitlock, P. A. (1986) *Monte Carlo methods. Volume I: Basics*, Wiley, New York, NY.

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

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 and System Safety*, 95, 550-564.

Rao, K.D., Kushwaha, H.S., Verma, A.K., Srividya A. (2007) Quantification of epistemic and aleatory uncertainties in level-1 probabilistic safety assessment studies. *Reliability Engineering and System Safety*, 92, 947-956.

USNRC (2009) *Guidance on the Treatment of Uncertainties Associated with PRAs in Risk-Informed Decision Making.* NUREG-1855, US Nuclear Regulatory Commission, Washington, DC.

Zadeh, L.A (1965) Fuzzy Sets, *Information and Control*, 8, 338-353.