HAL
open science

Line sampling and Fuzzy Interval Analysis for the propagation of aleatory and epistemic uncertainties in risk models

Elisa Ferrario, Nicola Pedroni, Enrico Zio

To cite this version:

Elisa Ferrario, Nicola Pedroni, Enrico Zio. Line sampling and Fuzzy Interval Analysis for the propagation of aleatory and epistemic uncertainties in risk models. ESREL 2013, Sep 2013, Amsterdam, Netherlands. pp.1-8. hal-00838667

HAL Id: hal-00838667

https://centralesupelec.hal.science/hal-00838667

Submitted on 26 Jun 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Line sampling and Fuzzy Interval Analysis for the propagation of aleatory and epistemic uncertainties in risk models

E. Ferrario, N. Pedroni, E. Zio
Chair on Systems Science and the Energetic Challenge, European Foundation for New Energy - Electricité de France, at École Centrale Paris - Supelec, France

E. Zio
Department of Energy, Politecnico di Milano, Italy

ABSTRACT: In this paper, an advanced Monte Carlo (MC) simulation method, namely Line Sampling (LS), is considered in combination with Fuzzy Interval Analysis (FIA) for improving the sampling efficiency in the hierarchical propagation of hybrid (probabilistic and possibilistic) uncertainties through a model for the risk-based design of a flood protection dike. The proposed approach is compared to that of standard MC sampling showing superior accuracy and higher computational efficiency.

1 INTRODUCTION

In this paper, we consider a "level-2" framework of two hierarchical levels of uncertainty (Limbourg & de Rocquigny 2010) representing aleatory (i.e. random) events whose stochastic laws of occurrence are described by models containing parameters (e.g. probabilities, failure rates, …) that are epistemically-uncertain, i.e. known with poor precision due to lack of knowledge and information.

We use probability distributions to describe aleatory uncertainty on the occurrence of the random events and possibility distributions to represent the epistemic uncertainty in the parameters of the (aleatory) probability distributions (Baudrit et al. 2006). The propagation of this hybrid (probabilistic and possibilistic) uncertainty representation can be performed by combining (Baudrit et al. 2008) the standard Monte Carlo (MC) technique (Kalos & Withlock 1986, Zio 2012) with the extension principle of fuzzy set theory (Zadeh 1965): this is done by (i) Fuzzy Interval Analysis (FIA) to process the epistemic uncertainty described by possibility distributions and (ii) repeated MC sampling of the random variables to process the aleatory uncertainty (Baudrit et al. 2008).

However, this approach can require considerable and often prohibitive computational efforts for two reasons: (i) a very large number of random realizations must be sampled by MC in order to perform an accurate uncertainty propagation, i.e. to perform a deep exploration of the (typically high-dimensional) space of the uncertain parameters of the model (Schueller 2009); (ii) the intervals computations of FIA require multiple evaluations of the model for each simulated random sample (Baudrit et al. 2006).

Efficient sampling methods would, then, be useful for reducing the number of random samples drawn and the associated computational effort. In this respect, techniques like Importance Sampling (IS), Stratified Sampling, Latin Hypercube Sampling (LHS), Subset Simulation (SS) and the Cross-Entropy (CE) method have been widely used in risk assessment problems (Zio 2012).

In the present paper, Line Sampling (LS) is implemented for improving the sampling efficiency (Koutsourelakis et al. 2004, Zio & Pedroni 2009, 2010). The basic idea of LS is to employ lines instead of random points in order to explore the (typically) high-dimensional space of the uncertain parameters of the model under analysis. An "important direction" is optimally determined in the multi-dimensional space of the uncertain parameters; then, the (multi-dimensional) uncertainty propagation problem is "decomposed" into a number of conditional, "one-dimensional" uncertainty propagation problems that are solved along such direction (Koutsourelakis et al. 2004, Zio & Pedroni 2009, 2010, Zio 2012).

The novel approach combining LS and FIA (called hereafter as "LS-FIA" approach for brevity) is here applied to hierarchically propagate hybrid uncertainties through a risk model for the design of a flood protection dike (Limbourg & de Rocquigny 2010). To the best of the author knowledge, it is the first time that the LS and FIA techniques are combined to hierarchically propagate hybrid uncertainties in risk assessment problems. The results are

compared to those obtained by a "classical" approach combining FIA and standard MCS (called hereafter as "MC-FIA" approach for brevity).

The remainder of the paper is organized as follows. In Section 2, the representation of aleatory (probabilistic) and epistemic (possibilistic) uncertainties in a "level-2" framework is described in detail; in Section 3, the main concepts underlying the hybrid LS-FIA approach are briefly outlined; in Section 4, the case study concerning the risk-based design of a flood protection dike is presented; in Section 5, the results of the joint hierarchical propagation of hybrid uncertainties through the flood risk model of Section 4 are reported, commented and compared to those produced by the standard MC-FIA approach; in Section 6, some conclusions are drawn. Finally, in Appendices A and B, the operative steps of the MC-FIA and LS-FIA approaches, respectively, are described in details.

2 "LEVEL-2" REPRESENTATION FRAMEWORK OF ALEATORY AND EPISTEMIC UNCERTAINTIES

We consider the (failure) behavior of a (possibly safety-critical) system described by a mathematical model $f(Y_1, Y_2, ..., Y_n)$, whose output Z is a function of n (input) parameters $\{Y_j : j = 1, 2, \ldots, n\}$. In risk assessment problems (like the one considered in this paper), the output Z is a physical quantity *critical* for safety reasons (e.g. the water level of a river close to a residential area, the temperature of the fuel cladding in a nuclear reactor, …).

Due to imprecise knowledge and limited information on the system, uncertainty is always present in the values of the system model input parameters/variables: this results into variability in the model output Z.

In this paper, we consider the system failure probability $P(\Omega)$, expressed as the probability that Z exceeds a safety threshold z^*, i.e. $P(\Omega) = P[Z > z^*]$ (Limbour & de Rocquigny 2010).

In all generality, we consider the uncertain (input) variables $\{Y_j : j = 1, 2, \ldots, n\}$, described by the Probability Distribution Functions (PDFs) $\{p_{\theta_j}^{Y_j}(y_j) : j = 1, 2, ..., n\}$ with epistemically-uncertain parameters $\boldsymbol{\theta}_j = \{\theta_{j,1}, \theta_{j,2}, ..., \theta_{j,P_j}\}$, $j = 1, 2, \ldots, n$, represented by possibility distributions $\boldsymbol{\pi}^{\theta_j}(\boldsymbol{\theta}_j) = \left\{\pi^{\theta_{j,1}}(\theta_{j,1}), \pi^{\theta_{j,2}}(\theta_{j,2}), ..., \pi^{\theta_{j,m}}(\theta_{j,m}), ..., \pi^{\theta_{j,P_j}}(\theta_{j,P_j})\right\}$.

For clarification by way of example, we may consider the generic uncertain variable Y_j described by a Gumbel PDF with parameters η (location parameter) and δ (scale parameter), i.e. $Y_j \sim Gum(\eta, \delta) = Gum(\boldsymbol{\theta}_j) = Gum(\theta_{j,1}, \theta_{j,2})$. Parameter $\delta = \theta_{j,2}$ (i.e. the scale parameter) is a fixed pointwise value ($\delta = \theta_{j,2}$ = 400), whereas parameter $\eta = \theta_{j,1}$ (i.e. the location parameter) is epistemically uncertain. The only information available on $\eta = \theta_{j,1}$ is that it is defined on interval $[a, b]$ = [900, 1300] and its most likely value is c = 1100. This limited state of knowledge about $\eta = \theta_{j,1}$ can be represented by a triangular possibility distribution $\pi^{\eta}(\eta)$ with core c = 1100 and support $[a, b]$ = [900, 1300] (Figure 1, left) (Baudrit & Dubois 2006).

Given the possibility function $\pi^{\eta}(\eta)$ of $\eta = \theta_{j,1}$, we can define its α-cut sets $A_{\alpha}^{\eta} = \{\eta : \pi^{\eta}(\eta) \geq \alpha\}$, for $0 \leq \alpha \leq 1$. For example, $A_{0.5}^{\eta}$ = [1000, 1200] is the set of η-values for which the possibility function is greater than or equal to 0.5 (dashed segment in Figure 1, left). Notice that the α-cut A_{α}^{η} of level α for parameter η can be interpreted as the $(1 - \alpha)$% Confidence Interval (CI) for η, i.e. the interval such that $P[\eta \in A_{\alpha}^{\eta}] \geq 1 - \alpha$ (Baudrit et al. 2006): for example, $A_{0.5}^{\eta}$ = [1000, 1200] is the 50% CI for η. In this view, the possibility distribution $\pi^{\eta}(\eta)$ can be interpreted as a set of *nested* CIs for parameter η (Baudrit & Dubois 2006).

For each possibility (respectively, confidence) level α (respectively, $1 - \alpha$) in [0, 1], a bundle of Cumulative Distribution Functions (CDFs) for Y_j can be constructed by letting parameter η range within the corresponding α-cut set A_{α}^{η}. This family of CDFs (of level α) is bounded above and below by the upper and lower CDFs, $\overline{F}_{\alpha}^{Y_j}(y_j)$ and $\underline{F}_{\alpha}^{Y_j}(y_j)$, defined as $\overline{F}_{\alpha}^{Y_j} = \sup_{\eta \in A_{\alpha}^{\eta}} \left\{F_{\eta}^{Y_j}(y_j)\right\}$ and $\underline{F}_{\alpha}^{Y_j} = \inf_{\eta \in A_{\alpha}^{\eta}} \left\{F_{\eta}^{Y_j}(y_j)\right\}$, respectively. Since $\pi^{\eta}(\eta)$ can be interpreted as a set of nested CIs for parameter η (see above), it can be argued that the α-cuts of $\pi^{\eta}(\eta)$ induce also a *set* of *nested pairs* of CDFs $\left\{\left(\underline{F}_{\alpha}^{Y_j}(y_j), \overline{F}_{\alpha}^{Y_j}(y_j)\right) : 0 \leq \alpha \leq 1\right\}$ which bound the "true" CDF $F^{Y_j}(y_j)$ of Y_j with confidence larger than or equal to $(1 - \alpha)$, i.e. $P[\underline{F}_{\alpha}^{Y_j}(y_j) \leq F^{Y_j}(y_j) \leq \overline{F}_{\alpha}^{Y_j}(y_j)] \geq 1 - \alpha$, with $0 \leq \alpha \leq 1$ (Baudrit et al. 2008). For illustration purposes, Figure 1, right shows the bounding upper and lower

CDFs of Y, $\overline{F}_{\alpha}^{Y_j}(y_j)$ and $\underline{F}_{\alpha}^{Y_j}(y_j)$, built in correspondence of the α-cuts of level α = 0 (solid lines), 0.5 (dashed lines) and 1 (dot-dashed line) of the possibility distribution $\pi^{\eta}(\eta)$ of parameter η.

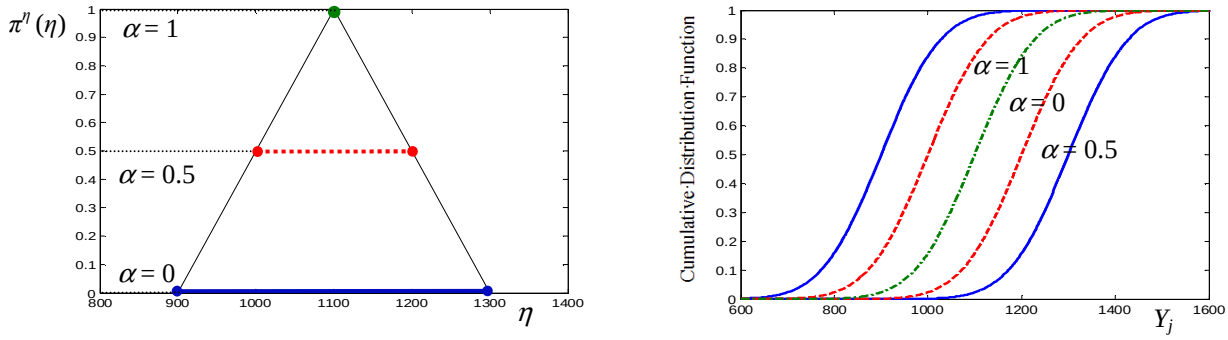

Figure 1. Left: Triangular possibility distribution of the parameter η of the Gumbel probability distribution of Y_j ~ Gum(η, 400); in evidence the α-cut of level α = 0, 0.5 and 1. Right: Upper and lower CDFs of Y_j built in correspondence of the α-cut of level α = 0, 0.5 and 1 for the parameter η.

3 LINE SAMPLING AND FUZZY INTERVAL ANALYSIS JOINT HIERARCHICAL PROPAGATION OF ALEATORY AND EPISTEMIC UNCERTAINTIES IN A "LEVEL-2" FRAMEWORK

The propagation of the hybrid (probabilistic and possibilistic) level-2" hierarchical uncertainty representation through the model $Z = f(Y_1, Y_2, ..., Y_n)$ can be performed by combining the standard Monte Carlo (MC) technique (Kalos & Withlock 1986, Zio 2013) with the extension principle of fuzzy set theory (Zadeh 1965), by means of the following main steps (Baudrit et al. 2008):

i. Fuzzy Interval Analysis (FIA) to process the epistemic uncertainty in the parameters $\boldsymbol{\theta}_j$ described by the possibility distributions $\boldsymbol{\pi}^{\theta_j}(\boldsymbol{\theta}_j)$, j = 1, 2, …, n;

ii. repeated *standard* MC sampling to process aleatory uncertainty described by the probability distributions $\{p_{\theta_j}^{Y_j}(y_j) : j = 1, 2, ..., n\}$.

Technical details about the operative steps of the hybrid MC-FIA approach are given in Appendix A.

The method produces a set of nested pairs of CDFs $\{(\underline{F}_{\alpha}^{Z}(z), \overline{F}_{\alpha}^{Z}(z)) : 0 \le \alpha \le 1\}$ for Z, that are used to estimate the system failure probability $P(\Omega) = P[Z > z^*] = 1 - P[Z < z^*] = 1 - F(z^*)$. Since the (hybrid) uncertainty in Z is represented by a set of nested pairs of CDFs, then $P(\Omega)$ is described by a possibility distribution $\pi^{P(\Omega)}$. Actually, for each possibility level α, two (upper and lower) estimates for the exceedance probability $P(\Omega)$, $\overline{P}(\Omega)_{\alpha}$ and $\underline{P}(\Omega)_{\alpha}$, are obtained as $\overline{P}(\Omega)_{\alpha} = 1 - \underline{F}_{\alpha}^{Z}(z^*)$ and $\underline{P}(\Omega)_{\alpha} = 1 - \overline{F}_{\alpha}^{Z}(z^*)$, respectively: then, the interval $[\underline{P}(\Omega)_{\alpha}, \overline{P}(\Omega)_{\alpha}]$ corresponds to the cut of level α of the possibility distribution $\pi^{P(\Omega)}$ of the exceedance probability $P(\Omega)$.

In this paper, step ii. of the procedure above (i.e. *standard* random sampling) is replaced by the adoption of Line Sampling (LS). To the authors knowledge, this is the first time that an accelerated sampling technique is introduced in the scheme of hybrid uncertainty propagation, to improve sampling efficiency.

As mentioned in the Introduction, LS has been originally developed for complex structural reliability problems (Koutsourelakis et al. 2004) and is based on the sampling of directed *lines* (instead of random *points*) for exploring the multi-dimensional space of the uncertain model parameters (Koutsourelakis et al. 2004, Zio & Pedroni 2009, 2010). In extreme synthesis, the problem of computing the system failure probability is transformed from the original "physical" space into the so-called "standard normal space", where each random variable is represented by an independent, central-unit Gaussian distribution. In the transformed space, a *unit* vector $\boldsymbol{\gamma}$ (hereafter also called "important direction") is determined, pointing towards the failure domain Ω of interest. The problem of computing the system failure probability in the original multi-dimensional space is then reduced to a number of conditional one-dimensional problems, which are readily solved in the standard normal space along the "important direction" by using the standard normal cumulative distribution function (Koutsourelakis et al. 2004, Zio & Pedroni 2009, 2010). The operative steps of the procedure are not given here for the sake of brevity; the interested reader is referred to Appendix B for technical details.

Several applications have shown that the method performs better than standard Monte Carlo Simula-

tion. Theoretically, if the boundaries of the failure domain of interest are not too rough (i.e. almost linear) and the "important direction" is almost perpendicular to them, the variance of the failure probability estimator could be reduced to zero (Koutsourelakis et al. 2004).

4 A CASE STUDY OF FLOOD PROTECTION RISK-BASED DESIGN

We consider a protection dike in a residential area closely located to a river with potential risk of floods. Two issues are of concern for its design: (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 by flooding risk analysis framework accounting for uncertainty.

The model considered calculates the maximal water level of the river Z (i.e. the output variable of the model) as a function of several parameters (i.e. the input variables of the model) (Limbourg & de Rocquigny 2010):

$$Z = Z_V + \left(\frac{Q}{K_S * B * \sqrt{(Z_m - Z_V)/L}} \right)^{3/5} \quad (1)$$

where: $Y_1 = Q$ is the yearly maximal water discharge (m^3/s); $Y_2 = Z_m$ and $Y_3 = Z_v$ are the riverbed levels (m asl) at the upstream and downstream parts of the river under investigation, respectively; $Y_4 = K_s$ 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:* $B = 300\,\text{m}$, $L = 5000\,\text{m}$;
- *uncertain variables:* Q, Z_m, Z_v, K_s.

The uncertain variables are affected by aleatory and epistemic uncertainties. The aleatory part of the uncertainty is described by probability distributions of defined shape (e.g. normal, Gumbel etc.). The parameters of the probability distributions describing the aleatory uncertainty are themselves affected by epistemic uncertainty, represented in terms of possibility distributions.

In this work, we adopt the hybrid probabilistic and possibilistic representations already proposed in (Pedroni et al. 2013). For space limitations, we will not recall the representations considered that is not the objective of the present paper. The interested reader is referred to (Pedroni et al. 2013) for further details and explanation of the reasons underlying these choices.

The final goal of the case study application is to determine the failure probability $P(\Omega)$ of the dike. This requires that we evaluate the probability $P(\Omega)$ that the maximal water level of the river Z (the model output) exceeds a given threshold z^* (the dike level), i.e. $P(\Omega) = P(Z > z^*)$: in the present paper, z^* = 55.5 m (Limbourg & de Rocquigny 2010). As highlighted before, since the maximal water level of the river Z (the model output) is described by a set of nested pairs of CDFs $\left\{\left(\underline{F}_\alpha^Z(z), \overline{F}_\alpha^Z(z)\right): 0 \le \alpha \le 1\right\}$, then $P(\Omega)$ is represented by the possibility distribution $\pi^{P(\Omega)}$ (see Section 3).

5 RESULTS

Figure 2 shows the estimates of the possibility distributions $\pi^{P(\Omega)}$ of the failure probability of the dike, $P(\Omega) = P(Z > z^* = 55.5\,\text{m})$, produced by the hybrid LS – FIA approach (dashed lines) and by the standard MC – FIA approach (solid lines) with N_T = 40000 random samples drawn. It can be seen that the two methods provide *very similar* results: actually, the two curves overlap almost completely.

In order to provide a (rough) quantitative indication of the *precision* of the estimates provided by the two methods, the standard deviations $\overline{\sigma}_\alpha(\overline{P}(\Omega)_\alpha)$ of the upper bounds $\overline{P}(\Omega)_\alpha$ of the α-cuts of the possibility distributions $\pi^{P(\Omega)}$ of $P(\Omega)$ are computed; for illustration purposes, the values of $\overline{\sigma}_\alpha(\overline{P}(\Omega)_\alpha)$ are reported in Table 1 for α = 0.2, 0.4, 0.6 and 0.8. It is evident that the precision of the LS-FIA method is *consistently higher* than that of the MC-FIA: actually, the standard deviations of the failure probability estimates produced by the LS-FIA method are about 23-42 times *lower* than those of the MC-FIA approach.

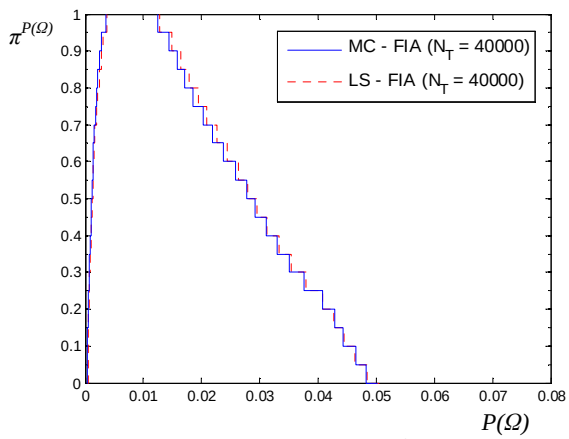

Figure 2: Possibility distributions $\pi^{P(\Omega)}$ of the failure probability P(Ω) of the dike, obtained by the MC – FIA (solid line) and LS – FIA (dashed line) approaches with N_T = 40000 random samples.

Table 1: Standard deviations $\bar{\sigma}_\alpha(\bar{P}(\Omega)_\alpha)$ of the upper bounds $\bar{P}(\Omega)_\alpha$ of the α-cuts of the possibility distribution $\pi^{P(\Omega)}$ for α = 0.2, 0.4, 0.6 and 0.8 (see Figure 2).

α	$\bar{\sigma}_\alpha(\bar{P}(\Omega)_\alpha)$ MC-FIA (N_T = 40000)	LS-FIA (N_T = 40000)
0.2	1.01E-03	4.31E-05
0.4	8.94E-04	3.04E-05
0.6	7.93E-04	2.24E-05
0.8	6.74E-04	1.61E-05

Figure 3 shows the estimates of the same quantities as before (see Figure 2), but obtained with N_T = 50 random samples (instead of 40000). This analysis is carried out to assess the performance of the two methods with a *very small* number of samples (and, thus, of system model evaluations): this is of paramount importance in practical cases in which the model requires several hours to run a single simulation. For the sake of comparison, Figure 3 reports also the possibility distribution produced by the standard MC–FIA approach with N_T = 40000 random samples (solid line): since this number of samples is large enough for precisely estimating the failure probability of the dike in the present case, this possibility distribution is ideally taken as the "real" one and it is considered a "reference" benchmark solution in the comparisons.

Notice that the result from LS-FIA (dashed line) is *very close* to this reference solution (solid line): actually, the two possibility distributions overlap almost completely; however, this result is obtained by LS-FIA at a *much lower* computational effort (i.e. using N_T = 50 samples instead of 40000). On the contrary, MC-FIA with N_T = 50 samples is *not* able to produce *accurate* results: this is pictorially demonstrated by the fact that the corresponding possibility distribution (dot-dashed line) lies very far from the reference solution (solid line). In addition, note that as before the precision of the LS-FIA method with N_T = 50 samples is *consistently higher* than that of the MC-FIA with the same number of samples: actually, the standard deviations of the failure probability estimates produced by LS-FIA are about 33-62 times *lower* than those of MC-FIA (Table 2).

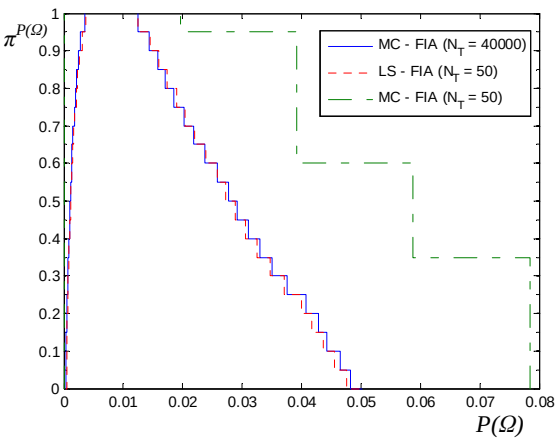

Figure 3: Possibility distributions $\pi^{P(\Omega)}$ of the failure probability P(Ω) of the dike, obtained by the MC – FIA (dot-dashed line) and LS – FIA (dashed line) approaches with N_T = 50 random samples; the result produced by the MC – FIA approach with N_T = 40000 samples (solid line) is also reported.

Table 2: Standard deviations $\bar{\sigma}_\alpha(\bar{P}(\Omega)_\alpha)$ of the upper bounds $\bar{P}(\Omega)_\alpha$ of the α-cuts of the possibility distribution $\pi^{P(\Omega)}$ for α = 0.2, 0.4, 0.6 and 0.8 (see Figure 3).

α	$\bar{\sigma}_\alpha(\bar{P}(\Omega)_\alpha)$ MC-FIA (N_T = 50)	LS-FIA (N_T = 50)
0.2	3.80E-02	1.15E-03
0.4	3.33E-02	8.23E-04
0.6	3.33E-02	6.14E-04
0.8	2.75E-02	4.43E-04

In addition to the precision of the failure probability estimate, also the *computational time* associated to the simulation method has to be taken into account. Table 3 reports the computational time t_{comp} associated to the simulations performed: it can be seen that the computational cost of the LS – FIA approach is about three times *higher* than that of the MC–FIA approach. Thus, in order to compare the *overall computational efficiency* of the two simulation methods (i.e. in order to take into account *both* the precision of the estimates produced *and* the computational time required), a synthetic indicator, namely Figure Of Merit (FOM), is considered. It is defined as $\mathrm{FOM} = 1/\sigma^2(P(\Omega)) \cdot t_{comp}$, where t_{comp} is the computational time required by the simulation method and $\sigma^2(P(\Omega))$ is the variance of the failure probability estimate. Notice that the *higher* is the value of the *index*, the *higher* is the *efficiency* of the method. Table 4 summarizes the values of the FOM for the MC-FIA and LS-FIA approaches computed using the values of the standard deviations $\bar{\sigma}_\alpha(\bar{P}(\Omega)_\alpha)$ given in Figures 2 and 3, right, and the values of the computational time t_{comp} reported in Table 3. It is evident that although the computational time of LS-FIA is higher than that of MC-FIA, its overall computational efficiency is still *consistently higher*: actually, the FOM of LS-FIA is 117-565 times larger than that of MC-FIA in the case of N_T = 40000, whereas it is 370-1305 times larger in the case of N_T = 50.

Table 3: Computational time t_{comp} [s] of the MC – FIA and LS – FIA approaches carried out with N_T = 40000 and N_T = 50 random samples.

	MC - FIA		LS - FIA	
Number of simulations, N_T	40000	50	40000	50
Computational time t_{comp} [s] on a Intel® Core™2 Duo CPU E7600 @ 3.06 and 3.07 GHz	31998	42	99181	124

Table 4: Values of the FOM characterizing the MC-FIA and LS-FIA approaches employing N_T = 40000 and 50 random samples

	FOM	
α	MC-FIA (N_T = 40000)	LS-FIA (N_T = 40000)
0.2	30.63	5427.71
0.4	39.10	10909.98
0.6	49.69	20094.42
0.8	68.79	38897.32
α	MC-FIA (N_T = 50)	LS-FIA (N_T = 50)
0.2	16.48	6097.93
0.4	21.47	11906.34
0.6	21.47	21391.51
0.8	31.48	41093.28

6 CONCLUSIONS

A combined Line Sampling (LS) and Fuzzy Interval Analysis (FIA) approach for the joint hierarchical propagation of hybrid aleatory (probabilistic) and epistemic (possibilistic) uncertainties has been proposed and exemplified with reference to a model for the design of a flood protection dike. In the exemplification, the uncertainty propagation task has been finalized to the estimation of the failure probability of the dike. The results obtained have been compared with those produced by a standard Monte Carlo (MC) and Fuzzy Interval Analysis (FIA) approach.

It has been shown that LS-FIA outperforms MC-FIA in terms of both *accuracy* and *precision* of the failure probability estimates. In addition, although the computational *time* associated to the LS-FIA approach is generally higher than that of MC-FIA, the *overall computational efficiency* (quantified by the FOM) is still consistently superior.

The outstanding performance of the LS method in combination with the FIA approach makes it a rather attractive tool for the propagation of "level-2" hybrid probabilistic-possibilistic uncertainties in risk assessment problems.

REFERENCES

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.

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

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

Koutsourelakis, P.S. Pradlwarter, H.J. Schueller, G.I. 2004. Reliability of structures in high dimensions, Part I: algorithms and application. Probabilistic Engineering Mechanics19: 409–17.

Limbourg, P. & 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.

Pedroni, N. Zio, E. Ferrario, E. Pasanisi, A. Couplet M. 2013. Hierarchical propagation of probabilistic and non-probabilistic uncertainty in the parameters of a risk model. Comput Struct, 10.1016/j.compstruc.2013.02.003.

Schueller, G.I. 2009. Efficient Monte Carlo simulation procedures in structural uncertainty and reliability analysis - recent advances. Journal of Structural Engineering and Mechanics 32(1): 1–20.

Zadeh, L.A. 1965. Fuzzy Sets. Information and Control 8: 338-353.

Zio, E. 2012. The Monte Carlo Simulation Method for System Reliability and Risk Analysis. London: Springer Series in Reliability Engineering.

Pedroni, N. & Zio, E. 2009. Functional failure analysis of a thermal–hydraulic passive system by means of Line Sampling. Reliability Engineering and System Safety 94: 1764-1781.

Zio, E. Pedroni, N. 2010. An optimized Line Sampling method for the estimation of the failure probability of nuclear passive systems. Reliability Engineering and System Safety 95(12): 1300–1313.

APPENDIX A: MONTE CARLO AND FUZZY INTERVAL ANALYSIS APPROACH

The main steps of the hybrid MC-FIA approach are the following (Baudrit et al. 2008):

1. set α = 0 (outer loop processing epistemic uncertainty by fuzzy interval analysis);
2. select the α-cuts $A_\alpha^{\theta_{j,1}}, A_\alpha^{\theta_{j,2}}, ..., A_\alpha^{\theta_{j,P_j}}$, j = 1, 2, …, n, of the possibility distributions $\boldsymbol{\pi}^{\boldsymbol{\theta}_j}(\boldsymbol{\theta}_j)$ = $\left\{\pi^{\theta_{j,1}}(\theta_{j,1}), \pi^{\theta_{j,2}}(\theta_{j,2}), ..., \pi^{\theta_{j,P_j}}(\theta_{j,P_j})\right\}$ of the parameters $\boldsymbol{\theta}_j$, j = 1, 2, …, n;
3. sample N_T random intervals $[\underline{y}_{j,\alpha}^{i_a}, \overline{y}_{j,\alpha}^{i_a}]$, $i_a = 1, 2, ..., N_T$, j = 1, 2, …, n, of the "probabilistic" variables Y_j, j = 1, 2, …, n, letting parameters $\boldsymbol{\theta}_j$ range within the corresponding α-cuts $A_\alpha^{\theta_{j,1}}, A_\alpha^{\theta_{j,2}}, ..., A_\alpha^{\theta_{j,P_j}}$, j = 1, 2, …, n (found at step 2. above) (inner loop processing aleatory uncertainty by standard MC simulation);
4. find the smallest and largest values of $Z = f(Y_1, Y_2, ..., Y_n)$, denoted by $\underline{z}_\alpha^{i_a}$ and $\overline{z}_\alpha^{i_a}$, respectively, letting variables Y_j range within the intervals $[\underline{y}_{j,\alpha}^{i_a}, \overline{y}_{j,\alpha}^{i_a}]$, $i_a = 1, 2, ..., N_T$, $j = 1, 2, ..., n$ (found at step 3. above);

5. take the values $\underline{z}_{\alpha}^{i_a}$ and $\overline{z}_{\alpha}^{i_a}$ found in 4. above as the lower and upper limits of the N_T α-cuts A_{α}^{Z,i_a} of Z, $i_a = 1, 2, ..., N_T$. A probability mass $m\left(A_{\alpha}^{Z,i_a}\right) = 1/N_T$ is associated at each α-cut A_{α}^{Z,i_a}, $i_a = 1, 2, ..., N_T$;
6. calculate the upper and lower CDFs $\overline{F}_{\alpha}^{Z}(z)$ and $\underline{F}_{\alpha}^{Z}(z)$ of level α as $\overline{F}_{\alpha}^{Z}(z) = \sum_{\inf\{A_{\alpha}^{Z,i_a}\} \le z} m(A_{\alpha}^{Z,i_a})$ and $\underline{F}_{\alpha}^{Y}(A) = \sum_{\sup\{A_{\alpha}^{Z,i_a}\} \le z} m(A_{\alpha}^{Z,i_a})$, respectively.

if $\alpha < 1$, then set $\alpha = \alpha + \Delta\alpha$ (e.g. $\Delta\alpha = 0.05$ in this paper) and return to step 2. above; otherwise, stop the algorithm.

APPENDIX B: LINE SAMPLING AND FUZZY INTERVAL ANALYSIS APPROACH

The operative steps for the propagation of hybrid probabilistic and possibilistic uncertainty in a "level-2" framework by LS – FIA approach are the following (Koutsourelakis et al. 2004, Zio & Pedroni 2009, 2010):

1. Determine the unit important direction $\boldsymbol{\gamma} = \{\gamma_1, \gamma_2, ..., \gamma_n\}$ that points towards the failure domain, Ω, of interest. The important direction is computed as the normalized "center of mass" of the failure domain of interest applying Markov chain to generate failure samples.
2. Sample in the normal space a matrix $\{\boldsymbol{\Psi}_j^q\}$, with q lines, $q = 1, 2, \cdots, N_T$, and j columns, $j = 1, 2, \cdots, n$, of random values of the variables Y_j, $j = 1, \cdots, n$; N_T is the number of simulations. Each component of the matrix $\{\boldsymbol{\Psi}_j^q\}$ is associated with an independent central unit Gaussian distribution.
3. Project each line of the random sample matrix $\{\boldsymbol{\Psi}_j^q\}$, $q = 1, 2, \cdots, N_T$, $j = 1, 2, \cdots, n$, onto the straight line passing through the origin O of the standard normal space and perpendicular to $\boldsymbol{\gamma}$ in order to obtain the matrix $\{\boldsymbol{\Psi}_j^{q,\perp}\}$:

$$\boldsymbol{\Psi}_j^{q,\perp} = \boldsymbol{\Psi}_j^q - \left\langle \boldsymbol{\gamma}, \boldsymbol{\Psi}_j^q \right\rangle \boldsymbol{\gamma}, \qquad \text{(A.1)}$$

where $q = 1, 2, \cdots, N_T$, $j = 1, 2, \cdots, n$

In (A.1) $\boldsymbol{\Psi}_j^q$, $q = 1, 2, \cdots, N_T$, $j = 1, 2, \cdots, n$, denotes a random realization of the input variables in the standard normal space of dimension n and $\left\langle \boldsymbol{\gamma}, \boldsymbol{\Psi}_j^q \right\rangle$ is the scalar product between $\boldsymbol{\gamma}$ and $\boldsymbol{\Psi}_j^q$, $q = 1, 2, \cdots, N_T$, $j = 1, 2, \cdots, n$. It is worth noting that since the standard Gaussian space is isotropic, the matrix $\boldsymbol{\Psi}_j^{q,\perp}$ is also standard normally distributed.
4. Define the sample matrix $\tilde{\boldsymbol{\Psi}}_j^q$, $q = 1, 2, \cdots, N_T$, $j = 1, 2, \cdots, n$, as a sum of a deterministic multiple of $\boldsymbol{\gamma}$ and the matrix $\boldsymbol{\Psi}_j^{q,\perp}$ in (A.1), i.e.

$$\tilde{\boldsymbol{\Psi}}_j^q = c^q \boldsymbol{\gamma} + \boldsymbol{\Psi}_j^{q,\perp}, \qquad \text{(A.2)}$$

where $q = 1, 2, \cdots, N_T$, $j = 1, 2, \cdots, n$ and c^q is a real number in [-∞, +∞]. Again, it is worth noting that since the standard Gaussian space is isotropic, the matrix $\tilde{\boldsymbol{\Psi}}_j^q$ is also standard normally distributed.
5. Moving along the straight line passing through $\{\boldsymbol{\Psi}_j^q\}$, $q = 1, 2, \cdots, N_T$, $j = 1, 2, \cdots, n$, and parallel to $\boldsymbol{\gamma}$, select three different values c_i^q, $i = 1, 2, 3$, for c^q and calculate the corresponding sample points $\tilde{\boldsymbol{\Psi}}_{j,i}^q$, $q = 1, 2, \cdots, N_T$, $j = 1, 2, \cdots, n$, $i = 1, 2, 3$, according to (A.2).
6. Compute the standard normal cumulative values $\boldsymbol{\Psi}_{j,i}^{Cum,q}$, $q = 1, 2, \cdots, N_T$, $j = 1, 2, \cdots, n$, $i = 1, 2, 3$, in correspondence of the sample points $\tilde{\boldsymbol{\Psi}}_{j,i}^q$ $q = 1, 2, \cdots, N_T$, $j = 1, 2, \cdots, n$, $i = 1, 2, 3$.
7. Set $\alpha = 0$ (outer loop processing epistemic uncertainty).
8. Estimate N_T lower and upper conditional "one-dimensional" failure probability estimates $\underline{P}^{1D,q}(\Omega)_{\alpha}$ and $\overline{P}^{1D,q}(\Omega)_{\alpha}$, $q = 1, 2, \cdots, N_T$, by the following steps:
   a. Select the α-cuts $A_{\alpha}^{\theta_{j,1}}, A_{\alpha}^{\theta_{j,2}}, ..., A_{\alpha}^{\theta_{j,P_j}}$ of the possibility distributions $\boldsymbol{\pi}^{\boldsymbol{\theta}_j}(\boldsymbol{\theta}_j) = \left\{ \pi^{\theta_{j,1}}(\theta_{j,1}), \pi^{\theta_{j,2}}(\theta_{j,2}), ..., \pi^{\theta_{j,P_j}}(\theta_{j,P_j}) \right\}$ of the parameters $\boldsymbol{\theta}_j = \{\theta_{j,1}, \theta_{j,2}, ..., \theta_{j,P_j}\}$, of the "probabilistic" variables $Y_1, Y_2, ..., Y_n$, as intervals of possible values $\left[\underline{\theta}_{j,\alpha}, \overline{\theta}_{j,\alpha}\right] = \left\{ \left[\underline{\theta}_{j,1,\alpha}, \overline{\theta}_{j,1,\alpha}\right], \left[\underline{\theta}_{j,2,\alpha}, \overline{\theta}_{j,2,\alpha}\right], ..., \left[\underline{\theta}_{j,P_j,\alpha}, \overline{\theta}_{j,P_j,\alpha}\right] \right\}$, $j = 1, 2, \cdots, n$;
   b. Set $q = 1$ (inner loop processing aleatory uncertainty);
   c. Compute the $q-th$ random intervals $\left[\underline{y}_{j,i,\alpha}^q, \overline{y}_{j,i,\alpha}^q\right]$, $j = 1, 2, \cdots, n$, $i = 1, 2, 3$, in the phys-

ical space, of the "probabilistic" variables Y_j, $j=1,2,\cdots,n$, corresponding to the α-cuts $\left\lfloor \boldsymbol{\theta}_{j,\alpha}, \bar{\boldsymbol{\theta}}_{j,\alpha} \right\rfloor = \left\{ \left\lfloor \underline{\theta}_{j,1,\alpha}, \bar{\theta}_{j,1,\alpha} \right\rfloor, \left\lfloor \underline{\theta}_{j,2,\alpha}, \bar{\theta}_{j,2,\alpha} \right\rfloor, ..., \left\lfloor \underline{\theta}_{j,P_j,\alpha}, \bar{\theta}_{j,P_j,\alpha} \right\rfloor \right\}$ (found at step 8.a. above) and to the $q-th$ random matrix $\{\boldsymbol{\Psi}_{1,i}^{Cum,q}, \boldsymbol{\Psi}_{2,i}^{Cum,q}, ..., \boldsymbol{\Psi}_{n,i}^{Cum,q}\}$ $i=1,2,3$, (generated at step 6. above). In particular, the $q-th$ random interval $\left[\underline{y}_{j,i,\alpha}^{q}, \bar{y}_{j,i,\alpha}^{q} \right]$ for Y_j, $j=1,2,\cdots,n$, is calculated by $\underline{y}_{j,i,\alpha}^{q} = \inf_{\theta_j \in [\underline{\boldsymbol{\theta}}_{j,\alpha}, \bar{\boldsymbol{\theta}}_{j,\alpha}]} \left[F_{\boldsymbol{\theta}_j}^{Y_j} \right]^{-1} \left(\boldsymbol{\Psi}_{j,i}^{Cum,q} \right)$ and $\bar{y}_{j,i,\alpha}^{q} = \sup_{\theta_j \in [\underline{\boldsymbol{\theta}}_{j,\alpha}, \bar{\boldsymbol{\theta}}_{j,\alpha}]} \left[F_{\boldsymbol{\theta}_j}^{Y_j} \right]^{-1} \left(\boldsymbol{\Psi}_{j,i}^{Cum,q} \right)$, $i=1,2,3$, where $\left[F_{\boldsymbol{\theta}_j}^{Y_j} \right]^{-1}$ is the inverse of the Cumulative Distribution Function (CDF) $F_{\boldsymbol{\theta}_j}^{Y_j}$ of $p_{\boldsymbol{\theta}_j}^{Y_j}$;

d. Calculate the smallest and largest values of $Z = f(Y_1, Y_2, ..., Y_j, ..., Y_n)$, $i=1,2,3$, denoted by $\underline{z}_{i,\alpha}^{q}$ and $\bar{z}_{i,\alpha}^{q}$ respectively, letting variables Y_j range within the intervals $\left[\underline{y}_{j,i,\alpha}^{q}, \bar{y}_{j,i,\alpha}^{q} \right]$, $j=1,2,...,n$, $i=1,2,3$; in particular, $\underline{z}_{i,\alpha}^{q} = \inf_{j, Y_j \in [\underline{y}_{j,i,\alpha}^{q}, \bar{y}_{j,i,\alpha}^{q}]} f(Y_1, Y_2, ..., Y_j, ..., Y_n)$ and $\bar{z}_{i,\alpha}^{q} = \sup_{j, Y_j \in [\underline{y}_{j,i,\alpha}^{q}, \bar{y}_{j,i,\alpha}^{q}]} f(Y_1, Y_2, ..., Y_j, ..., Y_n)$, $i=1,2,3$.

e. Normalize the values $\underline{z}_{i,\alpha}^{q}$ and $\bar{z}_{i,\alpha}^{q}$, $i=1,2,3$, found in step 8.d. above with respect to the threshold of interest, $z*$, computing the performance functions $\underline{g}_{i,\alpha}^{q}$ and $\bar{g}_{i,\alpha}^{q}$, respectively, as $\underline{g}_{i,\alpha}^{q} = \frac{\underline{z}_{i,\alpha}^{q}}{z*} - 1$ and $\bar{g}_{i,\alpha}^{q} = \frac{\bar{z}_{i,\alpha}^{q}}{z*} - 1$. Positive values of $\underline{g}_{i,\alpha}^{q}$ and $\bar{g}_{i,\alpha}^{q}$ indicate that the lower and upper values $\underline{z}_{i,\alpha}^{q}$ and $\bar{z}_{i,\alpha}^{q}$, $i=1,2,3$, respectively, exceed the threshold $z*$.

f. Fit the points $[c_i^q, \underline{g}_{i,\alpha}^{q}]$ and $[c_i^q, \bar{g}_{i,\alpha}^{q}]$, $i=1,2,3$, by means of first- or second-order polynomials and determine its roots $[\underline{c}_{\alpha}^{q}, \bar{c}_{\alpha}^{q}]$. The values $[\underline{c}_{\alpha}^{q}, \bar{c}_{\alpha}^{q}]$ represent the intersection between the limit state function $\underline{g}_{\alpha}^{q}(\underline{z}_{i,\alpha}^{q}, z*) = 0$ or $\bar{g}_{\alpha}^{q}(\bar{z}_{i,\alpha}^{q}, z*) = 0$, $i=1,2,3$, and the straight line passing through $\boldsymbol{\Psi}_j^q$ and parallel to $\boldsymbol{\gamma}$.

g. Calculate the lower and upper conditional "one-dimensional" failure probabilities estimate, $\underline{P}^{1D,q}(\Omega)_\alpha$ and $\bar{P}^{1D,q}(\Omega)_\alpha$, associated to each random samples $\boldsymbol{\Psi}_{j,i}^q$ $q=1,2,\cdots,N_T$, $j=1,2,...,n$, $i=1,2,3$ for the α-cut under evaluation, as:

$$\underline{P}^{1D,q}(\Omega)_\alpha = P[N(0,1) > \bar{c}_\alpha^q] = 1 - P[N(0,1) \le \bar{c}_\alpha^q] = 1 - \Phi(\bar{c}_\alpha^q) = \Phi(-\bar{c}_\alpha^q) \quad \text{(A.3)}$$

$$\bar{P}^{1D,q}(\Omega)_\alpha = P[N(0,1) > \underline{c}_\alpha^q] = 1 - P[N(0,1) \le \underline{c}_\alpha^q] = 1 - \Phi(\underline{c}_\alpha^q) = \Phi(-\underline{c}_\alpha^q) \quad \text{(A.4)}$$

where $\Phi(\cdot)$ denotes the standard normal cumulative distribution function.

h. If $q \ne N_T$, then set $q = q+1$ and return to step 8.c. above; otherwise go to step 9. below.

9. Compute the lower and upper unbiased estimators $\underline{P}(\Omega)_\alpha$ and $\bar{P}(\Omega)_\alpha$ for the failure probability, $P(\Omega) = P(Z > z*)$, as the sample average of the lower and upper independent conditional "one-dimensional" failure probability estimates $\underline{P}^{1D,q}(\Omega)_\alpha$ and $\bar{P}^{1D,q}(\Omega)_\alpha$, $q=1,2,\cdots,N_T$, in (A.3) and (A.4), respectively:

$$\underline{P}(\Omega)_\alpha = (1/N_T) \sum_{q=1}^{N_T} \underline{P}^{1D,q}(\Omega)_\alpha \quad \text{(A.5)}$$

$$\bar{P}(\Omega)_\alpha = (1/N_T) \sum_{q=1}^{N_T} \bar{P}^{1D,q}(\Omega)_\alpha \quad \text{(A.6)}$$

The variance of the estimators (A.5) and (A.6) can then be written, respectively, as:

$\underline{\sigma}_\alpha^2 \left(\underline{P}(\Omega)_\alpha \right) = (1/N_T) * \underline{\sigma}_\alpha^2 \left(\underline{P}^{1D,q}(\Omega)_\alpha \right)$ and $\bar{\sigma}_\alpha^2 \left(\bar{P}(\Omega)_\alpha \right) = (1/N_T) * \bar{\sigma}_\alpha^2 \left(\bar{P}^{1D,q}(\Omega)_\alpha \right)$, where

$$\underline{\sigma}_\alpha^2 \left(\underline{P}^{1D,q}(\Omega)_\alpha \right) = 1/(N_T - 1) \sum_{q=1}^{N_T} \left(\underline{P}^{1D,q}(\Omega)_\alpha - \underline{P}(\Omega)_\alpha \right)^2$$

and

$$\bar{\sigma}_\alpha^2 \left(\bar{P}^{1D,q}(\Omega)_\alpha \right) = 1/(N_T - 1) \sum_{q=1}^{N_T} \left(\bar{P}^{1D,q}(\Omega)_\alpha - \bar{P}(\Omega)_\alpha \right)^2$$

represent the sample variance of the independent conditional "one-dimensional" failure probability estimates.

10. if $\alpha \ne 1$, then set $\alpha = \alpha + \Delta\alpha$ (e.g. $\Delta\alpha = 0.05$) and return to step 8. above; otherwise, stop the algorithm.

The possibility distribution $\pi^{P(\Omega)}$ of the probability that the output $Z = f(Y_1, Y_2, ..., Y_n)$ exceeds a given threshold $z*$ is constructed as the collection of the values $\underline{P}(\Omega)_\alpha$ and $\bar{P}(\Omega)_\alpha$, $\alpha = 0, 0.05, ..., 0.95, 1$, found at step 8.g. above (in other words, $\pi^{P(\Omega)}$ is defined by all its α-cut intervals $[\underline{P}(\Omega)_\alpha, \bar{P}(\Omega)_\alpha]$).