

An Approach to Deal with Aleatory and Epistemic Uncertainty within the Same Framework: Case Study in Risk Assessment

Palash Dutta
Dept. of Mathematics
Dibrugarh University
Dibrugarh,-786004, India

ABSTRACT

Risk assessment is an important and popular aid in the decision making process. The aim of risk assessment is to estimate the severity and likelihood of harm to human health from exposure to a substance or activity that under plausible circumstances can cause to human health. In risk assessment, it is most important to know the nature of all available information, data or model parameters. More often, it is seen that available information model parameters, data are usually tainted with aleatory and epistemic uncertainty or both type of uncertainty. When some model parameters are affected by aleatory uncertainty and other some parameters are affected by epistemic uncertainty, how far computation of the risk is concern, one can either transform all the uncertainties to one type of format or need for joint propagation of uncertainties. In this paper, an effort has been made to combine probability distributions, normal fuzzy numbers and generalized interval valued fuzzy numbers (IVFNs) within the same framework.

KEYWORDS

Aleatory & Epistemic Uncertainty, Fuzzy Set, Generalized Fuzzy Number, Interval Valued Fuzzy Numbers, Hybrid Method, Risk Assessment..

1. INTRODUCTION

The aspect of uncertainty is an important and integral to any risk assessment process. For any decision making process involving risk the modeling and quantification of the uncertainties is required. The uncertainties are basically two types viz., aleatory and epistemic uncertainty. Aleatory uncertainty arises due to inherent variability, natural stochasticity, environmental or structural variation across space or through time, manufacturing or genetic heterogeneity among components or individuals, and variety of others sources of randomness. On the other hand epistemic uncertainty arises due to the insufficient knowledge about the world, which includes small sample sizes, detection limits, imperfections in scientific understanding etc. The age old classical probability theory is enough to handle aleatory type uncertainty. On the other hand Fuzzy set theory can be used to handle epistemic uncertainty. However, in some situation it is not always possible for a membership function of the type $\mu : X \rightarrow [0,1]$ to precisely assign one point from $[0,1]$ so it is more realistic to assign interval value. According to Gehrke at al. (1996) many people believe that assigning an exact number to expert's opinion is too restrictive and the assignment of an interval valued is more realistic. In such situations interval valued fuzzy set (IVFS) comes into picture. IVFS was developed in the 1970's. In May, 1975 Sambuc presented in his doctoral research (thesis) the concept of IVFS named as ϕ -fuzzy set. After development of IVFVs,

different researchers have been studied this issue and applied in different areas. An IVFS is a set in which every element has degree of membership in the form of an interval. One can say, IVFS consist of two membership function, one is upper membership function (UMF) and other is lower membership function (LMF).

Different effort have been made by different researchers for joint propagation of aleatory and epistemic uncertainty in the same computation of risk viz., Guyonnet et al. (1999, 2003), Baudrit et al. (2003, 2005, 2006, 2008), kentel et al. (2004), Anoop et al. (2006, 2008), Li et al. (2007), Rao et al. (2008), Baraldi et al. (2008), Helton et al. (2008), Limbourg et al. (2010), Flage et al. (2010, 2011), Dutta et al. (2012), Pedroni et al. (2012, 2013). In all their effort it is seen that proposed approaches work when representation of some input parameters are probabilistic while some are normal fuzzy numbers.

In this paper, an effort has been made to combine probability distributions, normal fuzzy numbers and generalized interval valued fuzzy numbers (IVFNs) within the same framework.

2. UNCERTAINTY MODELING TECHNIQUES:

2.1 Probability Theory:

Probability theory frequently used in uncertainty analysis. If parameters used in prescribed models are random in nature and followed well define distribution, then probabilistic methods are most suitable and well accepted approach for risk assessment.

A random variable is a variable in a study in which subjects are randomly selected. Let X be a discrete random variable.

A probability mass function is a function such that

$$(i) f(x_i) \geq 0, \quad (ii) \sum_{i=1}^n f(x_i) = 1 \quad (iii) f(x_i) = p(x = x_i)$$

The cumulative distribution function of a discrete random variable X, denoted as F(x) is

$$F(x) = P(X \leq x) = \sum_{x \leq x_i} f(x_i)$$

Let X be a continuous random variable. A probability density function of X is a non-negative function f, which satisfies

$$P(X \in B) = \int_B f(x) dx$$

for every subset B of the real line.

As X must assume some value, f must satisfy

$$P(X \in (-\infty, \infty)) = \int_{-\infty}^{\infty} f(x) dx = 1$$

This means the entire area under the graph of the PDF must be equal to unit.

In particular, the probability that the value of X falls within an interval $[a, b]$ is

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

The CDF of a continuous random variable X is

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(x)dx$$

2.2 Fuzzy Set Theory:

Environmental/human health risk assessment is an important aid in any decision-making process in order to minimize the effects of human activities on the environment. Unfortunately, usually environmental data tends to be vague and imprecise, so uncertainty is associated with any study related with these kinds of data. Fuzzy set theory provides a way to characterize the imprecisely defined variables, define relationships between variables based on expert human knowledge and use them to compute results. In this section, some necessary backgrounds and notions (Dutta et al., 2011a) of fuzzy set theory that will be required in the sequel are reviewed.

2.2.1: Let X be a universal set. Then the fuzzy subset A of X is defined by its membership function

$$\mu_A : X \rightarrow [0, 1]$$

Which assign a real number $\mu_A(x)$ in the interval $[0, 1]$, to each element $x \in A$, where the value of $\mu_A(x)$ at x shows the grade of membership of x in A .

2.2.2: Given a fuzzy set A in X and any real number $\alpha \in [0, 1]$. Then the α -cut or α -level or cut worthy set of A , denoted by ${}^\alpha A$ is the crisp set

$${}^\alpha A = \{x \in X : \mu_A(x) \geq \alpha\}$$

The strong α cut, denoted by ${}^{\alpha+}A$ is the crisp set

$${}^{\alpha+}A = \{x \in X : \mu_A(x) > \alpha\}$$

2.2.3: The support of a fuzzy set A defined on X is a crisp set defined as

$$\text{Supp}(A) = \{x \in X : \mu_A(x) > 0\}$$

2.2.4: The height of a fuzzy set A , denoted by $h(A)$ is the largest membership grade obtain by any element in the set and it is denoted as

$$h(A) = \sup_{x \in X} \mu_A(x)$$

2.2.5 An interval valued fuzzy set A defined in the universe of discourse X is represented by

$$A = \{(x, [\mu_A^L(x), \mu_A^U(x)]: x \in X)\}$$

Where $0 \leq \mu_A^L(x) \leq \mu_A^U(x) \leq 1$ and the membership grade $\bar{\mu}_A(x)$ of elements of x to the interval valued fuzzy set A is represented by an interval $[\mu_A^L(x), \mu_A^U(x)]$ (i.e., $\bar{\mu}_A(x) = [\mu_A^L(x), \mu_A^U(x)]$).

2.2.6 If an interval valued fuzzy set A satisfies the following properties

- A is normal
- A is defined in a closed bounded interval
- A is convex set

Then A is called an interval valued fuzzy number.

2.2.7 α -cut of IVFN:

Zeng & Shi (2005), Zeng & Li (2006) and Zeng et al. (2007) introduced and investigated the use of α -cut of IVFNs and how to extend arbitrary operations to such sets using interval α -cut rather than a single number. A generalization of α -cut of fuzzy set is

$$A^\alpha = \{x | \mu_A^L(x) \geq \alpha, \mu_A^U(x) \geq \alpha\}$$

3. PROPOSED HYBRID APPROACH

Consider a model

$$M = g(P_1, P_2, \dots, P_m, Q_1, Q_2, \dots, Q_r, F_1, F_2, \dots, F_n)$$

which is a function of parameters. Suppose P_1, P_2, \dots, P_m are m parameters presented by probabilistic distributions while Q_1, Q_2, \dots, Q_r are r parameters presented by normal fuzzy numbers and F_1, F_2, \dots, F_n are n parameters presented by generalized interval valued fuzzy numbers (IVFNs).

The approach is explained below:

Step1: Consider all normal fuzzy numbers Q_1, Q_2, \dots, Q_r as well as upper membership functions (upper fuzzy numbers) $F_1^u, F_2^u, \dots, F_n^u$ of generalized interval valued fuzzy numbers.

Step2: Then calculate α -cut for each fuzzy number. (First calculate $\alpha(=0)$ -cut for each fuzzy number (where $\alpha = [0, 1]$)). Then $r+n$ numbers of closed intervals (as α -cut gives closed intervals) i.e., $2(r+n)$ numbers of values will be obtained.

Step3: Generate m number of uniformly distributed random numbers from $[0, 1]$ and perform Monte Carlo simulation to obtain m numbers of random numbers by sampling probability distribution.

Step4: Assign all $2(r+n)$ values and m random numbers in the model M and calculate

$$M_1^{inf} = \text{Inf}(M) \text{ and } M_1^{sup} = \text{Sup}(M).$$

Step5: Repeat step1 to step4 for 5000 times

Step6: Plot cumulative distribution function (cdf) of

$$(M_1^{inf}, M_2^{inf}, \dots, M_{5000}^{inf}) \text{ and } (M_1^{sup}, M_2^{sup}, \dots, M_{5000}^{sup}).$$

Step7: Consider other α levels to calculate α -cut of each fuzzy numbers.

Step8: Repeat step 2 to step 6.

If proceed in this way a family of cdfs will be obtained.

Step9: Next, consider all normal fuzzy numbers Q_1, Q_2, \dots, Q_r as well as lower membership functions $F_1^l, F_2^l, \dots, F_n^l$ of generalized interval valued fuzzy numbers with heights w_1, w_2, \dots, w_n respectively.

Step10: Repeat step2 to step8. In step7 it should be noted that $\alpha = [0, w]$, where $w = \min(w_1, w_2, \dots, w_n, 1)$. Then we shall have another family of cdfs.

From these family of cdfs, membership functions (generalized interval valued fuzzy numbers) at different fractiles can be generated. First family of cdfs will produce UMF and later family of cdfs will give LMF with height w of the resulting generalized interval valued fuzzy number generated at different fractiles.

4. CASE STUDY

To demonstrate and make use of the proposed approach a hypothetical case study for non-cancer risk assessment is presented here. As due to the discharge of produce water into the sea a lot of organic and inorganic pollutants (however, in this example we consider only the heavy metal arsenic (As) because of its toxicity and high concentration in produced water.) release into the water and which are harmful to the aquatic organism. Therefore human being may be affected by ingestion of such contaminated aquatic organism. An evaluation is necessary to determine the possible impact such substances may have on human health and ecology. For this purpose, risk assessment is performed to quantify the potential detriment to human and evaluate the effectiveness of proposed remediation measures.

The general form of a comprehensive food chain risk assessment model as provided by EPA, 2001 is follows

$$CDI = \frac{C_f \times FIR \times FR \times EF \times ED \times CF}{BW \times AT} \dots\dots\dots(1)$$

Where CID = Chronic daily intake (mg/kg-day), FIR = fish ingestion rate (g/day), FR = fraction of fish from contaminated source, EF = exposure frequency (day/year), ED = exposure duration (years), CF = conversion factor (= 10⁻⁹), BW = body weight (kg), AT = averaging time (days) and Cf = chemical concentration of fish tissue (mg/kg). The chemical concentration in fish tissue (Cf) can be computed as

$$C_f = PEC \times BCF \dots\dots\dots(2)$$

Where PEC = predicted environmental concentration (mg/l) and BCF is the chemical bioaccumulation factor in fish (l/kg).

The non-cancer risk model for fish ingestion is expressed as:

$$Risk_{non-cancer} = \frac{CDI}{Rfd} \dots\dots\dots(3)$$

Where, Rfd is the reference dose.

In this case study, representation of the parameters predicted environmental concentration (PEC) is considered as normal fuzzy number while chemical bioaccumulation factors (BCF) is considered to be generalized interval valued fuzzy number (IVFN), body weight (BW) is taken as normal probability distribution and other parameters are taken to be constant. Values of the parameters for the calculation of non-cancer risk are given in the table 1.

Table 1: Parameter values used in the risk assessment

Parameter	Units	Type of Variable	Value/distribution
Average Time (AT)	Days	Constant	25550
Body Weight (BW)	Kg	Probabilistic	Normal(70,5)
Exposure Duration (ED)	Years	Constant	30
Exposure frequency (EF)	Days/year	Constant	350
Fraction of contaminated Fish (FR)	-	Constant	0.5
Fish Ingestion Rate (FIR)	g/day	Constant	170
Conversion Factor (CF)	-	Constant	1E-09
PEC for As	$\mu g / L$	Fuzzy	[4 ,5 ,6]
BCF for As	L/kg	IVFN	[35, 45, 55] UMF [40, 45, 50;0.8] LMF
Oral Rfd for As	mg/(kg.day)	Constant	3.0E-04

In this case study, representation of the parameters predicted environmental concentration (PEC) is considered as normal fuzzy number while chemical bioaccumulation factors (BCF) is considered to be generalized interval valued fuzzy number (IVFN), body weight (BW) is taken as normal probability distribution and other parameters are taken to be constant. Values of the parameters for the calculation of non-cancer risk are given in the table 1. The uncertain variables BW , PEC and BCF are plotted in figures 1, 2 and 3 respectively.

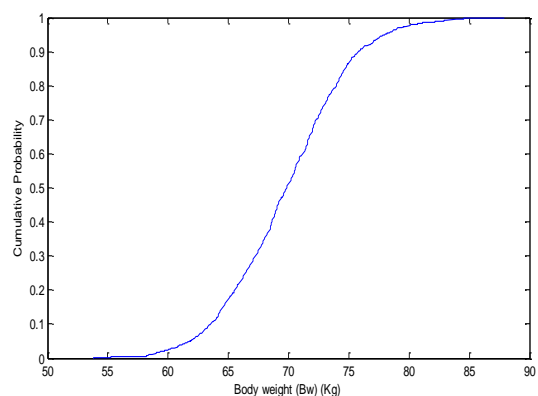


Figure1: Uncertain variable Bw (Kg)

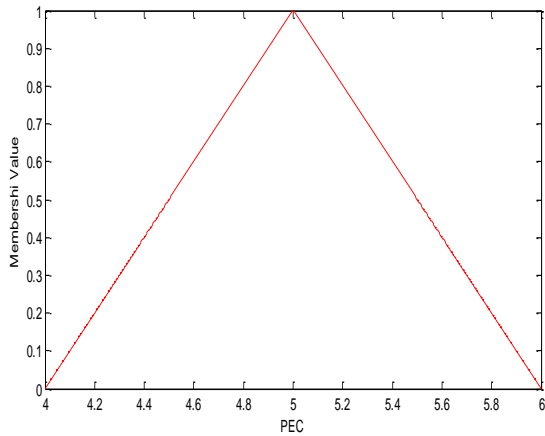


Figure 2: Uncertain variable PEC ($\mu\text{g/L}$)

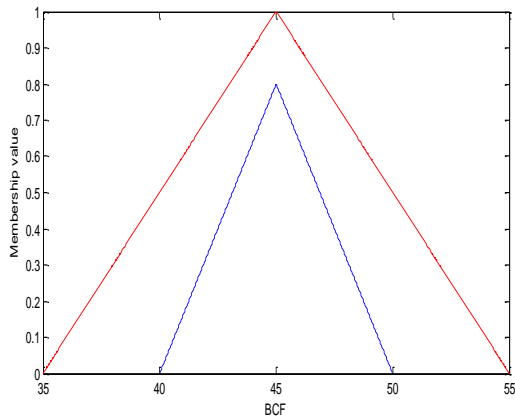


Figure 3: Uncertain variable BCF (L/Kg)

The result of the non-cancer human health risk assessment is performed using our proposed hybrid approach and which is depicted in figure (4).

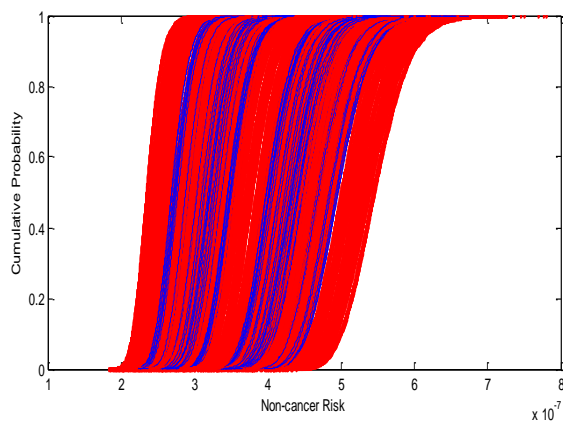


Figure 4: Cumulative distribution functions of non cancer risk for different α values

The result of the risk assessment is obtained in the form of family of Cdfs (basically two family of cdfs, one in red coloured and another in blue coloured) at different α -values. Red coloured Cdfs are obtained for UMF and blue coloured Cdfs are obtained for LMF of the uncertain input parameter BCF. From these cdfs, risk at different fractiles (Maxwell et

al, 1998, Kentel et al, 2004 & Dutta et al, 2011b) can be calculated and which are obtained in the form of generalized interval valued fuzzy number. It is because any arithmetic operations between generalized fuzzy numbers and normal fuzzy numbers produces generalized fuzzy number.

For instance, at 95th fractile, non-cancer risk value lies in the interval valued fuzzy number {[2.63e-07, 42.24e-07, 6.214e-07] (UMF); [3.016e-07, 4.071e-07, 4.411e-07, 5.655e-07] (LMF)}. Similarly, at 85th fractile, risk values lie in the generalized fuzzy number {[2.515e-07, 4.03e-07, 5.929e-07] (UMF); [2.875e-07, 3.924e-07, 4.148e-07, 5.39e-07] (LMF)}. Their graphical representations of the resulting non cancer risk value at 95th and 85th fractiles are depicted below.

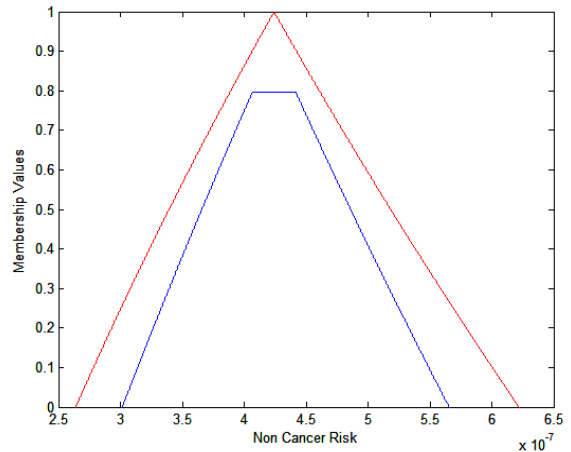


Figure 5: Membership function of risk at 95th fractile

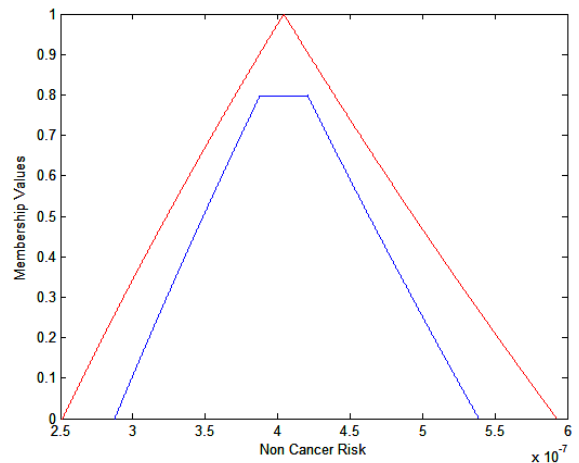


Figure 6: Membership function of risk at 85th fractile

5. CONCLUSION

In risk assessment, it is most important to know the nature of all available information, data or model parameters. More often, it is seen that available information is interpreted in probabilistic sense because probability theory is a very strong and well established mathematical tool to deal with aleatory uncertainty. However, not all available information, data or model parameters are affected by aleatory uncertainty (i.e., nature of the data, information or parameters are random) and can be handled by traditional probability theory. Imprecision may occur due to scarce or incomplete information or data, measurement error or data obtain from expert judgment or subjective interpretation of available data or information.

Thus, model parameters, data may be affected by epistemic uncertainty. Fuzzy set theory or possibility theory can be explored to handle this type of uncertainty (Dutta et al. 2012). Sometimes, it is also seen that some model parameters are affected by aleatory uncertainty and some parameters are affected by epistemic uncertainty, how far computation of the risk is concern, one can either transform all the uncertainties to one type of format or need for joint propagation of uncertainties. In this paper we have proposed a method to deal with such situation where some possibilistic distributions are considered as interval valued fuzzy numbers together with normal fuzzy numbers. To demonstrate and make use of the proposed hybrid method a hypothetical case study for non-cancer risk assessment is presented here. After performing risk assessment using our approach risk is obtained in the form of Cdfs and from which, membership functions of the risk are generated at different fractiles. The membership functions of risk at different fractiles are generalized interval valued fuzzy number since representation of at least one parameter is taken as generalized interval valued fuzzy number (IVFN). The lower membership function (LMF) of the generalized interval valued fuzzy number is trapezoidal type generalized fuzzy number, because any arithmetic operation of generalized fuzzy numbers (also generalized fuzzy number and normal fuzzy number) produces trapezoidal type generalized fuzzy number.

6. REFERENCES

- [1] Anoop M.B., Balaji Rao K., Gopalakrishnan S., Conversion of probabilistic information into fuzzy sets for engineering decision analysis. *Comp. and Struct.* 2006, 84(3–4): 141–155
- [2] Anoop M. B., Balaji Rao K., Lakshmanan N. Safety assessment of austenitic steel nuclear power plant pipelines against stress corrosion cracking in the presence of hybrid uncertainties. *Int. J. Pres.Vessels Piping.* (2008),85(4): 238–247
- [3] Baraldi, P. & Zio, E., A Combined Monte Carlo and Possibilistic Approach to Uncertainty Propagation in Event Tree Analysis. *Risk Analysis*, (2008.)28 (5): 1309-1326.
- [4] Baudrit C., Dubois D., Guyonnet D., Fargier H., Joint Treatment of imprecision and Randomness in Uncertainty Propagation, *Proc. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, Perugia, (2004) 873-880.
- [5] Baudrit C., Dubois D., Guyonnet D., Joint Propagation and Exploitation of Probabilistic and Possibilistic Information in Risk Assessment, *IEEE Transaction on Fuzzy Systems*, 14 (2006) 593-608.
- [6] Baudrit, C. & Dubois, D., Practical Representations of Incomplete Probabilistic Knowledge. *Computational Statistics & Data Analysis*, (2006.) 51 (1): 86-108.
- [7] Baudrit, C., Dubois, D., Perrot, N. Representing parametric probabilistic models tainted with imprecision. *Fuzzy Sets and System*, (2008) 159: 1913-1928.
- [8] Dubois D, Prade H, Sandri S., Onpossibility/probability transformations, (1993) In:RLowen,MRoubens (Eds.) *Fuzzy Logic*. (Dordrecht: Kluwer Academic Publishers).
- [9] Dutta P., Ali T., A Hybrid Method to Deal with Aleatory and Epistemic Uncertainty in Risk Assessment, *International Journal of Computer Applications*, (2012) 42: 37-44.
- [10] Dutta P., Ali T., Fuzzy Focal Elements in Dempster-Shafer Theory of Evidence: Case study in Risk Analysis, *International Journal of Computer application*, (2011b) 34: 46-53.
- [11] Dutta P., Boruah H., Ali T., Fuzzy Arithmetic with and without using α -cut method: A Comparative Study, *International Journal of Latest Trends in Computing.* (2011a) 2: 99-107.
- [12] EPA U.S., Risk Assessment Guidance for Superfund, Volume I: Human Health Evaluation Manual (Part E, Supplemental Guidance for Dermal Risk Assessment). Office of Emergency and Remedial Response, EPA/540/R/99/005, Interim, Review Draft. United States Environmental Protection Agency. September 2001.
- [13] Flage, R., Baraldi, P., Zio, E., Aven, T., Possibilityprobability transformation in comparing different approaches to the treatment of epistemic uncertainties in a fault tree analysis. In B. Ale, I.A. Papazoglu, E. Zio (Eds.), *Reliability, Risk and Safety - Proceedings of the ESREL, Conference, Rhodes, Greece* (2010).: 714-721.
- [14] Flage, R., Baraldi, P., Zio, E., Aven, T., Probabilistic and possibilistic treatment of epistemic uncertainties in fault tree analysis. (2011) Submitted to Risk analysis.
- [15] Gehrke M., Walker C., Walker E., Some comments on interval valued fuzzy sets, *Int. Jour. Intelligent Systems*, (1996) 11: 751-759.
- [16] Guyonnet D., Bourguine B., Dubois D., Fargier H., Côme B., Chilès J. P., Hybrid approach for addressing uncertainty in risk assessments, *Journal of Environmental Engineering*, (2003) 126: 68-78.
- [17] Haldar A, Reddy R K., A random-fuzzy analysis of existing structures. *Fuzzy Sets Syst.* (1992) 48:201–210
- [18] Helton J C, OberkampfWL., Alternative representations of epistemic uncertainty. *Reliab. Engng.Syst. Safety* (2004) 85:1–3
- [19] Kentel E., Aral M. M., Probalistic-fuzzy health risk modeling, *Stoch Envir Res and Risk Ass* (2004) 18: 324-338.
- [20] Li J., Huang G. H., Zeng G. M., Maqsood I., Huang Y. F., An integrated fuzzy-stochastic modeling approach for risk assessment of groundwater contamination, *Journal of Environmental Management*, (2007) 82: 173-188.
- [21] Limbourg, P. and de Rocquigny, E., Uncertainty analysis using evidence theory – confronting level-1and level-2 approaches with data availability and computational constraints, *Reliability Engineering andSystem Safety*, (2010) 95:550-564.
- [22] Maxwell R. M., Pelmulder S.D., Tompson A. F. B., Kastenber W. E., On the development of a new methodology for groundwater-driven health risk assessment, *Water Resources Res*, (1998) 34: 833–847.
- [23] Pedronia N., Zioa E., Ferrariob E., Pasanisid A., Couplet M., Propagation of aleatory and epistemic uncertainties in the model for the design of a flood protection dike, "PSAM 11 & ESREL (2012), Helsinki : Finland.

- [24] Pedronia N., Zioa E., Ferrariob E., Pasanisid A., Couplet M. Hierarchical propagation of probabilistic and non-probabilistic uncertainty in the parameters of a risk model, *Computers & Structures*, (2013): 1-15.
- [25] Rao, K.D., Kushwaha, H.S., Verma, A.K., Srividya A. Quantification of epistemic and aleatory uncertainties in level-1 probabilistic safety assessment studies. *Reliability Engineering and System Safety*, (2007) 92: 947-956.
- [26] Sambuc R., “Function Φ -Flous, Application a l’aide au Diagnostic en Pathologie Thyroïdienne”, These de DoctoratenMedicine,UniversityofMarseille (1975).
- [27] Zeng, W., Shi, Y. (2005), ‘Note on interval-valued fuzzy set’, *Lecture notes in computer science*, 3613, 20–25.
- [28] Zeng, W. & Li, H. (2006b), ‘Representation theorem of interval-valued fuzzy set’, *Uncertainty,Fuzziness and Knowledge-Based Systems* 14(3), 259–269.
- [29] Zeng, W., Li, H., Zhao, Y. & Yu, X. (2007), Extension Principle of Lattice-Valued Fuzzy Set, *in* ‘Fuzzy Systems and Knowledge Discovery, 2007. FSKD 2007. Fourth International Conference on’, Vol. 1.