WORKSHOP ON THE EVALUATION OF UNCERTAINTIES IN RELATION TO SEVERE ACCIDENTS AND LEVEL II PROBABILISTIC SAFETY ANALYSIS
CADARACHE, FRANCE, 7-9 NOVEMBER 2005

TITLE

Influence of Mathematical Modelling of Knowledge. Application to the Transfer of Radionuclides in the Environment

Authors :
Eric Chojnacki IRSN/DPAM/SEMIC
Catherine Mercat-Rommens IRSN/DEI/SESURE
Cédric Baudrit Université Paul Sabatier de Toulouse

IRSN : Institut de Radioprotection et de Sûreté Nucléaire

ABSTRACT

Although Monte-Carlo methods provide extremely flexible and powerful techniques for solving many of the uncertainty propagation problems encountered in safety analysis, these methods present two major drawbacks. Like most methods based on probability theory, Monte-Carlo methods need a lot of knowledge. Indeed to determine the probability law associated to each uncertain parameter, it is necessary to have collected a considerable amount of data or to make assumptions in the place of such empirical information. Moreover, to perform a Monte-Carlo simulation, it is also required to provide information about all the possible dependencies between the uncertain parameters. Unfortunately, in practice, such information is rarely fully available and the impact of the assumptions made to mitigate this lack of knowledge can deteriorate the relevance of the decision-making.

To overcome both kinds of limitations, the French Institute for Radiological Protection and Nuclear Safety (IRSN) in collaboration with BRGM, INERIS and University of Toulouse intends to experiment recent advances in Dempster Shafer theory. The purpose of this paper is to introduce our developments of this theory through the example of the transfer of a radionuclide in the environment. In particular, it will be shown that fuzzy numbers are a natural way to represent current state of radioecological knowledge because they can be considered equivalent to a family of probability distributions instead of a single one. Thus, the use of fuzzy numbers in uncertainty analysis allows to avoid the subjectivity which may exist in the choice of a single probability distribution. Afterwards, we will see how the Dempster-Shafer theory provides an unified theoretical framework allowing both the use of probability functions for stochastic uncertainties modelling and fuzzy numbers for epistemic uncertainties modelling.

It will also be shown that the stochastic independence assumption often taken by default in Monte-Carlo applications may lead to an artificial reduction of evaluated uncertainty ranges. On the contrary, the notion of epistemic independency, taken by default in fuzzy calculations, may lead to an artificial extension of evaluated uncertainty ranges. In this way, it appears that the epistemic independency is more related to an ignorance of stochastic dependencies than an assumption of stochastic independency.

The results obtained from our example with this uncertainty methodology can be extended to other realistic studies as far as the nature of monotonicity between uncertain variables and the response is known. Such kind of uncertainty methodologies seems to us the only possibility to warrant the robustness of Monte-Carlo simulation in case of incomplete knowledge.

Introduction

Taking into account uncertainties has become a major aspect of the estimation of pollutant transfer in the environment. The Monte-Carlo method which is based on probability theory requires to model uncertainties by the joint Probability Density Function (PDF) of all the uncertain parameters. To define this joint PDF, it is necessary to provide a (marginal) PDF for each uncertain parameter and also all the possible dependencies between the uncertain parameters. Unfortunately, such a knowledge is rarely available. To overcome this lack of knowledge in practical studies, a principle of minimal information is used. This principle consists for example to select the uniform distribution as PDF when only the uncertainty range of the parameter is known or to take an independence assumption between two uncertain parameters when no information is known about their dependence. However, this current practice does not respect the precautionary principle and may lead to an underestimation of the uncertainty. Indeed, the use of an uniform distribution between the minimum and the maximum values of an uncertain parameter leads to affect a slight probability to intervals such as [max – δx , max] for small δx values. In the same way, to assume stochastic independence between two uncertain parameters mean that the occurrence of both extreme values for these parameters has a very low probability. Therefore each one of this assumption on the uncertain parameters (marginal PDF choice and independence) may lead to an unjustified reduction of the final uncertainty of the response. A common practice to avoid such unwanted and artificial reduction of the estimated uncertainty, is to perform deterministic calculations with penalizing values for some uncertain parameters. Such a practice is in fact similar to interval calculation. However, this practice may lead to unrealistic extension of the response uncertainty and does not allow to distinguish the likelihood of values inside the uncertainty span.

In the first section of this paper, it will be shown that fuzzy numbers are a way to generalize interval calculation and the links between fuzzy numbers and possibility theory will be presented in the second section. Then, we will describe recent advances in uncertainty methodology based on the 'Dempster-Shafer' theory which allows to combine possibility and probability theories. In the last section, a simple

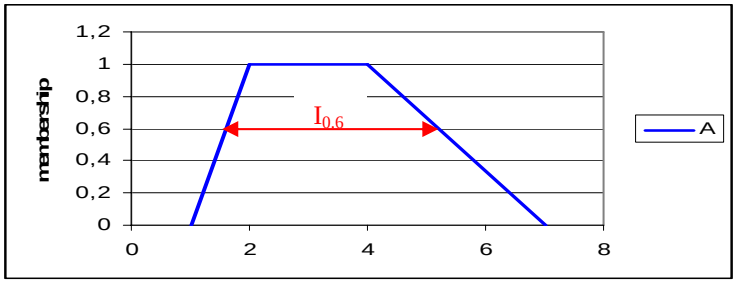
case of radionuclides transfer will allow us to see the benefits of the 'Dempster-Shafer' uncertainty methodology.

1. Fuzzy modelling[1] : an extension of interval calculation

The interval calculation consists into evaluating the interval of responses knowing the intervals of uncertain parameters. The scalar result obtained from a single calculation performed with fixed values for parameters is then replaced by an interval range containing all the results obtained when the parameters values are located within intervals.
For example, let us consider two parameters A and B with the following interval ranges : A=[1 , 7] and B=[2 , 4]. That means that the quantity A (resp. B) may take any value within [1 , 7] (resp. =[2 , 4]). The interval calculation for the result A+B is straightforward : A+B=[3 , 11]. In the same way, AB=[2 , 28], A/B=[1/4 , 7/2] etc.

A fuzzy number is an interval for which it is given a membership value $\mu(x)$ for any real value x of this interval. The membership value is a real value between 0 and 1. For example, let us consider the trapezoidal fuzzy number A = (1, 2, 4, 7).

Trapezoidal fuzzy number A = (1, 2, 4, 7).

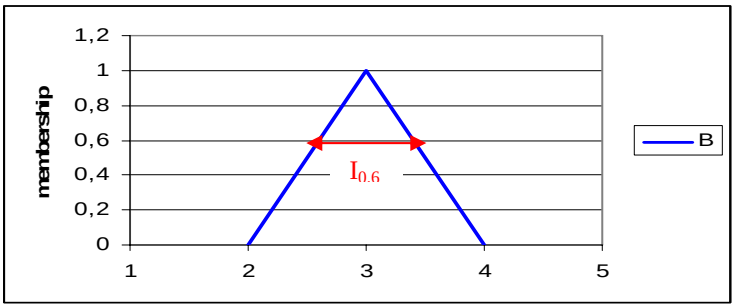


For each membership value α, it can be associated an interval I_α (called in fuzzy theory α-cut) : $I_\alpha = \{ x \mid \mu(x) \geq \alpha \}$. In this way, it appears that a fuzzy number is a set of nested intervals I_α : $I_1 \subset I_\alpha \subset I_0$. I_1 is called the core of the fuzzy number and I_0 its support. The membership value allows to distinguish the likelihood of the possible values in the support. A fuzzy number is therefore identical to a set of nested intervals weighted by their likelihood.

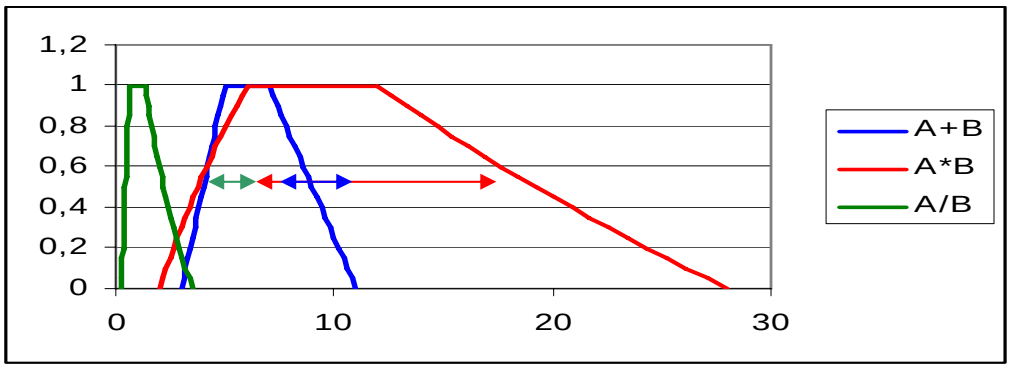
From the concept of fuzzy numbers, interval calculation can be easily generalized. Indeed, it is enough to perform a set of interval calculation for each α-cut. For each value of α, we consider all the intervals I_α associated to the uncertain parameters and in this way we obtain the interval I_α of the result. These intervals can now be used to generate the fuzzy number associated to the result. This set of interval calculation is called fuzzy arithmetic.

Let us illustrate the fuzzy arithmetic from an example.
Choose again for A the trapezoidal fuzzy number A = (1, 2, 4, 7) and for B the triangular fuzzy number B = (2, 3, 4)

Triangular fuzzy number B = (2, 3, 4).



Consider again the responses : A+B, A*B, A/B



Each interval I_α (as shown on the figure with α=0.6) of a response (here A+B, A*B and A/B) is obtained by a single interval calculation. The set of intervals I_α for a response defines the fuzzy number corresponding to this response.
The use of fuzzy numbers allows to take into account the likelihood of different values within the supports of uncertain parameters and constitutes in this way a generalization of interval calculation.

2. Links between fuzzy numbers and possibility theory[2]

A fuzzy number, as we have seen, is defined from a membership function which is a real function and therefore does not allow to measure the confidence in events which are sets. The possibility theory as the probability theory is a set-function which provide a measure of the confidence associated to an event.
By definition, a possibility measure must satisfy the three following axioms :

$\pi(\emptyset)=0$,
$\pi(E)= 1$ where E is the whole set,
$\pi(A\cup B)=\max(\pi(A), \pi(B))$ for any subsets A and B.

From a fuzzy number F, one can define a possibility measure π_F in a very simple way : for any set A, $\pi_F(A)=\max(\mu_F(x)\ ,\ x\in A)$ where μ_F is the membership function associated to the fuzzy number F. In this way, the possibility of an event A appears to be the maximal value of the membership in the set A. That provides a measure of the confidence in the event A knowing the information F.

The use of the possibility theory, based on fuzzy calculation, is consequently an improvement of the classical deterministic calculations, but may lead also to an

unrealistic extension of uncertainty estimations with respect to probabilistic calculations. Indeed, fuzzy calculation does not take into account the possible compensating effects due to the small likelihood to have several independent random variables at their extreme values.

3. Dempster-Shafer theory : an unified framework for possibility and probability theories[3]

The Dempster-Shafer theory is a generalization of the probability theory. The underlying idea of this theory is to extend the third Kolmogorov axiom of probability theory.
Let us remind the Kolmogorov axioms of a probability measure P :

$P(\emptyset)=0$,
$P(E)= 1$ where E is the whole set,
$P(A\cup B)= P(A)+P(B))$ for any subsets A and B such as $A\cap B= \emptyset$.

Thus, Shafer has defined a plausibility measure Pl with the axioms :

$Pl(\emptyset)=0$,
$Pl(E)= 1$ where E is the whole set,
$Max(Pl(A),Pl(B)) \leq Pl(A\cup B) \leq Pl(A)+Pl(B))$ for any subsets A and B.

From this definition of plausibility measures, possibility and probability measures appear to be particular plausibility measures. It can be shown that plausibility measures are defined from mass functions on subsets. For probability measures, the mass functions are based on singletons. Indeed, as a probability measure is additive, the probability of a set A is the sum of probabilities of each of its elements : the singletons. For possibility measures, the mass functions are based on nested intervals which are the intervals I_α.

Thus, the Dempster-Shafer theory provides a theoretical framework which allow to mix probabilistic calculations with fuzzy calculations (which are as we have already noticed an extension of classical interval calculations). In this way, it is possible to develop uncertainty methodologies which can benefit from the advantages of probabilistic and possibilistic calculations while avoiding their drawbacks.

Based on Dempster-Shafer theory, it is now possible to define an uncertainty methodology enough flexible to respect the real state of knowledge by relaxing the constraints inherent in the probabilistic methods :

– First, the use of possibility distributions, when and only when it is necessary, allows to relax assumptions currently made on the choice of the PDFs associated to some uncertain parameters. Actually, the use of fuzzy numbers and of their induced possibility distributions allows to model in a convenient way a family of PDFs and in this way it is not any more required to select a specific PDF when not enough information is known about the true distribution. For example a flat fuzzy number encodes all the PDFs which have the same support, a triangular fuzzy number encodes all the PDFs which have the same support and the same mode, a trapezoidal fuzzy number encodes the family of PDFs which have the same support and for which the mode is in the core. However, as probability theory is included in

‘Dempster-Shafer’ theory, specific PDFs can be used to model random variables, limiting the over-conservatism encountered in standard interval calculations.

– Secondly, this uncertainty methodology allows to model the possible ignorance of stochastic dependencies. In standard Monte-Carlo studies, the use of stochastic independency assumption taken by default when no knowledge is known about dependencies, may lead to an artificial reduction of uncertainty margins. Indeed, the stochastic independency between two uncertain variables means that it is unlikely to have simultaneously extreme values between random variables and lead to a compensating effect between uncertainty sources. On contrary, in possibilistic calculations, it is assumed that the information related to the uncertain variables have the same reliability or in an equivalent way that uncertain variables are epistemically independent. With this assumption, the uncertainties may cumulate themselves. In this way, it can be said that the epistemic independency is similar to an ignorance of the stochastic dependency. Indeed, the use of the epistemic independency as in the interval calculation, leads to cumulate uncertainties when no information is available about compensating effects. On the contrary, the use of stochastic independence assumptions, when it is possible, limits the over-conservatism of standard interval calculations.

The ‘Dempster-Shafer’ uncertainty methodology we have developed, consists in using PDF for an uncertain parameter when the knowledge is sufficient to determine it and to use a fuzzy number to model parameter uncertainty in other cases. The uncertainty derived by this methodology is a plausibility distribution. Indeed, as both probability and possibility distributions are used to model parameters uncertainty, the uncertainty attached to a result can no more be described by a probability or a possibility distribution, but is described by a plausibility. The knowledge of the plausibility distribution does not allow to know accurately the probability of events but only its lower and upper probabilities. Therefore the uncertainty attached to a result can be summarized by the data of its lower and upper probabilities, the difference between these probabilities coming from the lack of knowledge modelled by fuzzy numbers.

The uncertainty methodology which we propose, allows the analyst to cumulate the uncertainties when no knowledge is known so that our uncertainty estimations respect the precautionary principle and to benefit from compensating effects when stochastic independence is justified.

4. Application to a simple example of radionuclides transfer into the environment

We will consider a simplified radionuclide transfer model from maize ensilage to man through the consumption of milk. The absorbed dose D is calculated as the product of the maize ensilage activity A, the quantity of maize ensilage eaten by cow Q, the transfer factor from maize ensilage to milk F and the quantity of ingested milk I :

$$D = A*Q*F*I$$

For this example, we consider that the maize ensilage activity is a random variable, known from experimental measurement and the available information on the other quantities Q, F and I is only partial. We will assume that the knowledge relative to these quantities is as follows :

— the quantity of maize ensilage eaten by cow is very likely between 10 and 14 kg/day and cannot be out of the interval [4 , 35],

— the transfer factor from maize ensilage to milk F cannot be out of the interval [0.001 , 0.005] with 0.003 day/litre for the most likely value,

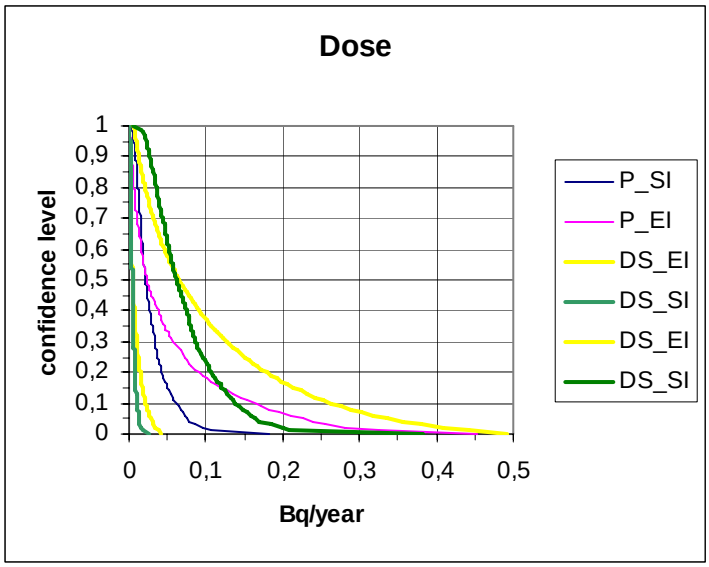
— the quantity of ingested milk I cannot be out of the interval [70 , 280] with 140 litre/year for the most likely value.

To evaluate the uncertainty attached to the absorbed dose, we will use both a standard probabilistic modelling and a Dempster-Shafer modelling. Each of these modellings will be performed with two independency assumptions : stochastic independence (P_SI and DS_SI methods, for respectively the Probabilistic and Dempster-Shafer modelling) and epistemic independence (P_EI and DS_EI methods).

Let us sum up in the following table, our choice of modelling for the uncertain parameters :

parameter	Probabilistic methods : P_SI and P_EI	Dempster-Shafer methods : DS_SI and DS_EI
A	random variable : lognormal distribution m=-5.76 , σ=0.58	random variable : lognormal distribution m=-5.76 , σ=0.58
Q	random variable : trapezoidal distribution (4, 10, 14, 35)	fuzzy variable : trapezoidal distribution (4, 10, 14, 35)
F	random variable : triangular distribution (0.001, 0.003, 0.005)	fuzzy variable : triangular distribution (0.001, 0.003, 0.005)
I	random variable : triangular distribution (70, 140, 280)	fuzzy variable : triangular distribution (70, 140, 280)

From these data, we performed Monte-Carlo simulations on random variables and also on fuzzy numbers with our 'Dempster-Shafer' method. The results of these numerical simulations allow to derive the Complementary Cumulative Density Function of the dose distribution when probabilistic methods are used and to derive a couple of CCDFs corresponding to the lower and upper CCDFs when the 'Dempster-Shafer' method is used.



The CCDF obtained with the probabilistic modelling is lying between the lower and upper CCDFs derived from the Dempster-Shafer modelling, as far as the same dependencies assumptions are taken. For example, with the stochastic independence assumption, the percentile 95% derived from a probabilistic modelling, is equal to 0.08 Bq/year (resp. 0.23 Bq/year with the epistemic independence) and is between [0.01, 0.16 Bq/year] (resp. between [0.03, 0.34 Bq/year] with the epistemic independence) when fuzzy numbers are used to avoid the choice of a specific PDF for some uncertain parameters.
These different curves show the importance of the assumptions related to the choice of marginal distributions (a factor ~10 on the percentile 95%) and their dependencies (a factor ~3 on the percentile 95%) on the uncertainty margins.

Conclusion

A classical argument to reject the use of probability theory to derive the uncertainty margins is that, in current applications, the available knowledge is not sufficient to know the PDFs associated to uncertain parameters and their possible dependencies. Consequently, the estimation of percentiles derived from Monte-Carlo simulations can be misleading.
Our method, based on Dempster-Shafer theory, allows an extension of classical Monte-Carlo simulation by relaxing assumptions relative to the choice of PDFs and their dependencies. The results provided by this method are consequently less precise (e.g. an interval instead of a scalar value for the percentiles) but are more reliable. Moreover, as it has been briefly explained, it is interesting to notice that this method can also be understood as an extension of classical deterministic interval calculations. Indeed, this method allows both to generalize interval calculation by the mean of fuzzy numbers and also by including random variables. Therefore, the obtained results are more accurate and discriminating than those derived from usual interval calculations.

References :

[1]Kazuo Tanaka
An introduction to fuzzy logic for practical applications
Springer 1997

[2]Dubois D., Nguyen H.T., Prade H.
Possibility theory, probability and fuzzy sets : Misunderstandings, bridges and gaps
In DuboisD., Prade H. Eds: Fundamentals of Fuzzy Sets, Kluwer, Boston, Mass, 343-438,2000

[3]Shafer G.
A Mathematical Theory of Evidence
Princeton – University Press 1976

[4]Baudrit C., Dubois D.
Comparing methods for joint objective and subjective uncertainty propagation with an example in risk assessment
4th International Symposium on Imprecise Probabilities and Their Applications,Pittsburgh, Pennsylvania, 2005