A hybrid approach for addressing uncertainty in risk assessments

Accepté pour publication dans *Journal of Environmental Engineering*

By : Dominique Guyonnet[1], Bernard Bourgine[1], Didier Dubois[2], Hélène Fargier[2], Bernard Côme[3] and Jean-Paul Chilès[1]

[1] : Envir. Spec., BRGM, BP 6009, 45060, Orléans Cédex 2, France
[2] : Math. Spec., Université Paul Sabatier, 31063 Toulouse, France
[3] : Envir. Spec., ANTEA, BP 6119, 45061 Orléans, Cédex 2, France

Key words : Risk assessment, Uncertainty, Monte Carlo, Fuzzy calculus

ABSTRACT : Uncertainty is a major aspect of the estimation, using models, of the risk of human exposure to pollutants. The Monte Carlo method, which applies probability theory to address model parameter uncertainty, relies on a statistical representation of available information. In recent years, the theory of possibilities has been proposed as an alternative approach to address model parameter uncertainty in situations where available information are insufficient to identify statistically representative probability distributions, due in particular to data scarcity. In practice, it may occur that certain model parameters can be reasonably represented by probability distributions, because there is sufficient data available to substantiate such distributions by statistical analysis, while others are better represented by fuzzy numbers (due to data scarcity). The question then arises as to how these two modes of representation of model parameter uncertainty can be combined for the purpose of estimating the risk of exposure. In this paper an approach (termed a hybrid approach) for achieving such a combination is proposed, and applied to the estimation of human exposure, via vegetable consumption, to cadmium present in the surficial soils of an industrial site located in the north of France. The application illustrates the potential of the proposed approach, which allows the uncertainty affecting model parameters to be represented in a fashion which is consistent with the information at hand.

INTRODUCTION

As risk assessments have become important aids in the decision-making process related to the management of sources of contamination, the issue of uncertainty with respect to model parameter values is of primary importance. Uncertainty affecting parameters in risk assessments can be of different nature (ambiguity, vagueness, imprecision, ignorance, etc.) and can be represented in various fashions (probability density functions, fuzzy numbers, uncertainty intervals, etc.). While different types of uncertainty may warrant different modes of uncertainty representation, the question arises as to how several modes of representation can be accommodated in the same estimation of risk ? Many researchers have addressed uncertainty using either one or the other of these modes of representation. For example Labieniec et al. (1997) used probability distribution functions to address uncertainty in the estimation of the risk of human exposure due to the presence of contaminated land. Prado et al. (1999) applied probability theory in risk assessments related to the underground disposal of nuclear waste. Dou et al. (1995), Bardossy et al. (1995), Freissinet et al. (1998), Cazemier (1999) present applications of fuzzy theory to environmental problems.

Comparatively few researchers have addressed the issue of combining different modes of representation of uncertainty in a same calculation (see for example Wonneberger et al., 1995). Yet this problem is of particular relevance in a risk assessment context, because in practice there is a very contrasted detail of information regarding the various parameters which influence the exposure to pollutants. This paper proposes a method (termed «hybrid») for combining probability distribution functions and fuzzy numbers in a same estimation of risk. The method is first explained, and then applied to estimate the risk of human exposure to cadmium present in the surficial soils of an industrial site located in the north of France. It should be noted that the proposed method is one among several possible alternatives, the relative advantages of which are currently being investigated. The primary purpose of this paper is to promote a methodology which is consistent with the information at hand. As shown below, the proposed methodology might also be consistent with a "reasonable" application of the precautionary principle.

THEORY

Although it is not the purpose of this paper to provide a detailed presentation of the theories of probability and possibility, for the sake of clarity some basic principles are presented below. For illustration purposes we consider a variable X which will be represented on the one hand using probabilities, and on the other hand using fuzzy theory. If X is considered as a random variable, and the probabilities that X should be less or equal than values x of X are known, a probability distribution function $F(x)$ can be defined as :

$$F(x) = P(X \leq x) \tag{1}$$

If variable X is continuous, i.e. it can take any value within a defined range, a probability density function $f(x)$ can be defined as :

$$f(x) = \frac{dF(x)}{dx} \tag{2}$$

These functions are illustrated graphically in Fig. 1(a), where a Gaussian probability distribution is assumed for variable X (in this case soil porosity). The distribution is defined by its mean (0.3) and its standard deviation (0.05). Ideally in true situations, the definition of such a distribution would rely on the collection of a sufficient number of porosity measurements, so that meaningful statistical moments could be derived by fitting the theoretical distribution to a histogram of measured data relative frequencies. When dealing with risk assessments, however, it often occurs that for some parameters there is insufficient data to perform such a frequential analysis. In recent years there has been a large amount of research, especially from the nuclear industry, regarding how probability distribution functions can be derived using expert judgement (see for example Bonano et al., 1990). So-called “a priori”, or subjective probability distribution functions can be defined using Bayesian probability theory (see for example Cullen and Frey, 1999). A fundamental postulate of this theory is that for any quantity, there exists a unique function which describes its probability distribution. In the estimation of risk, the probability distribution functions are sampled a large number of times, in a random fashion, using the Monte Carlo method (see for example Vose, 1996).

Some researchers (for example Wonneberger et al., 1995) have questioned the validity of deriving statistical moments or probability distribution functions from guesswork, even though based on expert knowledge. An alternative could be to consider values x of variable X as “possible”, rather than “probable”. According to the theory of possibilities (Zadeh ,1965, 1978 ; Dubois & Prade,1988), variable X can be represented by a “fuzzy number”, which describes the relationship between X and a membership function (noted μ). This function, the values of which are comprised between zero and one, describes the likelihood that variable X may take a certain value x. Figure 1(b) illustrates this concept for the same case as previously (soil porosity). Figure 1(b) informs us that it is considered most likely that porosity lies between 0.25 and 0.4, with no preference being expressed within this range, while values as low as 0.15 or as high as 0.5 cannot be totally excluded, and values outside this range are considered impossible. Possibility theory (see Dubois & Prade, 1988) provides the mathematical basis for performing calculations with fuzzy numbers.

Note that although a probability density function could be defined with exactly the same shape as the membership function in Fig. 1(b), a fundamental difference between a fuzzy number and a probability density function is that in the former case, the integral below the function need not be equal to unity. Compared with a probability density function, a fuzzy number is very “poor” in terms of intrinsic information. A probability density function defines a variable entirely, but applies best to systems which are “closed” (such as the throw of a dice). But the systems of interest in an environmental context are “open” (Oreskes et al., 1994), and possibility theory may be seen as a viable alternative for representing some of the uncertainty pertaining to such systems.

One question which was addressed in part by Guyonnet et al. (1999) is : “well what does it change in practice ?”. An important consequence in an environmental or health context is that the *a priori* assumption of probability distribution functions (noted hereafter PDF’s), without justification by available information, may lead to an unconservative estimation (minimisation) of risk. In order to understand this one must recall that the probability that two independent events, A and B, should occur simultaneously, is the product of the probabilities of both events.

Therefore, during Monte Carlo random sampling, scenarios that combine low probability parameter values have all the less chance of being selected. If a very large number of iterations is used, these scenarios will be realised, but with very low relative frequencies. When the results of the Monte Carlo analysis are compared with an acceptance criterion (for example a reference dose), for a certain level of probability (for example 95%), these scenarios will be eliminated because they fall within the 5% high outliers. Had model parameter uncertainty been considered in terms of possibilities rather than probabilities, these low-likelihood scenarios might not have been discarded, because fuzzy calculus does not transmit through multiplication the uncertainty of the parameter values onto that of the calculation result.

In the event that a given model involves some parameters which are justifiably represented by PDFs (in particular because there is sufficient data to substantiate these PDFs), while others are considered to be better represented by fuzzy numbers, a method should be derived to combine these two modes of representation of uncertainty in the estimation of risk. There are different ways by which this might be achieved. For example Wonneberger et al. (1995) suggest using possibility/probability transformations (Dubois & Prade, 1993, 1992). Another approach would be to treat both types of uncertainty within a single framework, using the "belief functions" proposed by Shafer (1976) (see also Smets and Kennes, 1994). While different alternatives exist, the approach described below (called the "hybrid approach") is believed to be more intuitively amenable to the practising environmental engineer. The respective advantages (and degrees of validity) of different alternatives to this problem are currently being investigated.

In order to illustrate the hybrid approach, we will consider the estimation of a dose resulting from the exposure of a human target to soil pollutants. This dose is calculated using a "model", M, which is a function of a certain number of parameters :

$$Dose = M(P_1, \ldots, P_n, F_1, \ldots, F_m), \tag{3}$$

where M = model; $P_1, \ldots, P_n$ = n model parameters each represented by a PDF; $F_1, \ldots, F_m$ = m model parameters each represented by a fuzzy number. Note that the model can also involve precise (i.e. "crisp") parameter values. Calculation of the dose is performed by combining the Monte Carlo random sampling technique, with the method of α-cuts (Dubois and Prade, 1988) for fuzzy calculus. An α-cut (see Fig. 2) is the ensemble of parameter values for which the likelihood of occurrence is greater or equal to α. In the application presented here, fuzzy calculus can be thought of as an interval analysis for different levels of likelihood. The combination procedure is summarised below, where steps have been indented to emphasise the iterative character of the calculations :

1. Generate n random numbers ($\chi_1, \ldots, \chi_n$) from a uniform distribution and sample the n PDF's to obtain a realisation of the n random variables : $p_1, \ldots p_n$ (Fig. 3A)
   2. Select a value α of the membership function (a level of likelihood).
      3. Calculate the *Inf* (smallest) and *Sup* (largest) values of $M(p_1, \ldots, p_n, F_1, \ldots, F_m)$, considering all values located within the α-cuts for each fuzzy number (see Fig. 3.B).
      4. Affect these *Inf* and *Sup* values to the lower and upper limits of the α-cut of $M(p_1, \ldots, p_n, F_1, \ldots, F_m)$.
   5. Return to step 2 and repeat steps 3 and 4 for another α-cut (note : α can be increased stepwise from 0 to 1 every 0.1 increments). The fuzzy result of $M(p_1, \ldots, p_n, F_1, \ldots,$

F_m) (the fuzzy dose) is obtained from the *Inf* and *Sup* values of $M(p_1, \ldots, p_n, F_1, \ldots, F_m)$ for each α-cut.

6. Return to step 1 to generate a new realisation of the random variables.

If steps 2 through 5 are repeated ω times, ω fuzzy doses are calculated (Fig. 3.C). For each value of the membership function (each value of α), the spread between the *Inf* and *Sup* values of the fuzzy results (see Fig. 3.C) is entirely a consequence of the Monte Carlo random sampling. It is therefore proposed to select the final *Inf* and *Sup* values of $M(P_1, \ldots, P_n, F_1, \ldots, F_m)$, for each value of α, by building a histogram of cumulative relative frequencies of the *Inf* and *Sup* values, and extracting the final *Inf* and *Sup* values for a certain level of probability. This is illustrated in Fig. 4. For each level of the histograms reproduce the spread of the *Inf* and *Sup* values. The final *Inf* and *Sup* values are taken such that there is a 5% probability of having values lower of higher respectively. This final fuzzy dose can be compared to an acceptance criterion, using the tools provided by possibility theory, as illustrated in the next section. Note that in the general case, step 4 of the procedure above can be performed using a minimisation and maximisation algorithm. However, if the model is an equation involving simple operations such as multiplication and substraction, the *Inf* and *Sup* values can be identified directly.

One important consequence of combining probability distributions with fuzzy numbers is that the net result is a fuzzy number : as the information conveyed by a fuzzy number is "poor" compared to a probability distribution, no mathematical procedure can compensate for this paucity and achieve the degree of system definition required by a probabilistic representation. The proposed hybrid approach takes advantage of the "rich" information provided by the PDF's, but retains the conservative character of fuzzy calculus to account for those parameters for which a representation by PDF's is not justified by available data. As will be shown in the closing section, if all parameters were assumed to be PDF's, the range of results would be narrowed and insufficient weight would be given to outlier scenarios which might have important implications in the decision-making process. On the other hand, if all parameters were represented by fuzzy numbers, despite the fact that some of them could be justifiably represented by PDF's, the range of results would be too conservative. Thus the proposed hybrid approach might hold some value in terms of consistency with a reasonable application of the precautionary principle.

APPLICATION TO AN INDUSTRIAL SITE

The proposed method is applied to a metallurgical industrial site located in the north of France. The surficial soils of this site are contaminated by a number of metallic pollutants among which cadmium, which have been deposited by the smoke emanating from a chimney located on the site. The quality of the surficial soils (upper ten centimeters) has been monitored in detail, and there is a relatively large amount of data available. These data are primarily total soil metal contents measured by atomic absorption spectrophotometry after extraction by fluorhydric acid. No selective extraction data (see for example Tessier et al., 1979) were available for this study. This detail is of importance since we are interested here in the exposure of a human target to cadmium in the soil via the consumption of vegetables. A large body of scientific evidence shows that there is poor correlation between the amount of cadmium absorbed by plants, and the total amount present in the soil (see for example Jopony and Young, 1993, Lorenz et al., 1997). Metal uptake by plants depends on a variety of factors among which metal speciation, plant specie, pH

(Dijkshoorn et al., 1983, Singh et al., 1995), redox conditions, humidity, temperature (Chang et al., 1987), competition with other metals (Smilde et al., 1992, Chaney et al., 1999), etc. The variety of these factors explains why it is difficult in practice to develop a model of metal uptake by plants based on total soil concentrations, and therefore preferable to have direct site-specific measurement of metal uptake by plants. This is fortunately the case for this industrial site : measurements of cadmium content were performed (Luttringer & de Cormis, 1979) for a limited number of vegetables (in particular leeks), grown in the immediate vicinity of the soil sampling points. Measurements were performed on the edible vegetable parts, after they had been washed to eliminate the metal fraction present in the dust at the surface of the leaves. Cadmium was analysed by atomic absorption after plant calcination and attack by chlorhydric and fluorhydric acid.

The first step in the analysis was to generate a map of soil cadmium content based on the measured data. Due to the relatively large number of surficial soil analyses (124), the spatial distribution of cadmium in the soils could be investigated using geostatistics (see Chilès and Delfiner, 1999). Geostatistics is a special branch of statistics which applies to data which display a spatial structure. Its use is not compulsory for the hybrid approach proposed in this paper, but served here to provide statistically representative estimators of soil cadmium concentrations outside measurement point locations. As could be expected, the data display a decrease in soil cadmium concentrations with increasing distance to the chimney (see Fig. 5 where Cd concentrations are in logarithm). In order to apply the classical geostatistical tools (the variogram and kriging), the data were first transformed into logarithm and then decomposed into a trend and a residual around this trend, according to :

$$\ln(\mathrm{Cd_s}) = \ln(\mathrm{Cd_s})_{\mathrm{T}} + R \tag{4}$$

where $\ln(\mathrm{Cd_s})$ = logarithm of measured soil cadmium concentrations; $\ln(\mathrm{Cd_s})_{\mathrm{T}}$ = predicted values of $\ln(\mathrm{Cd_s})$ according to the trend; R = residual around this trend. A correlation equation which describes the trend is :

$$\ln(\mathrm{Cd_s})_{\mathrm{T}} = -0.3 + 4.2\exp(-d/2) \tag{5}$$

where d = distance to the chimney. The spatial distribution of the residual is then examined using the variogram. As directional variograms did not reflect any anisotropy in the spatial distribution, an omnidirectional variogram was used. It was fitted with a linear variogram model and a nugget effect which accounts for microstructures and/or measurement errors (Fig. 6). The next step consisted in using this variogram model to interpolate the residual R, by kriging, which ensures that the interpolation is not biased (on average the kriging error is zero) and that it is optimal (the kriging variance is minimum). Noting the kriged residual as R_K, we obtain the soil cadmium concentration from :

$$\mathrm{Cd}^{*}_{\mathrm{sM}} = \exp\left(R_K + \ln(\mathrm{Cd_s})_{\mathrm{T}}\right) \tag{6}$$

where $\mathrm{Cd}^{*}_{\mathrm{sM}}$ is a median estimator of soil cadmium concentration since kriging of R yields a median estimator of R (i.e., the true value has a 50% probability of being lower than the kriged value). Results depicted in Fig. 7 show relatively high values of soil cadmium concentrations close to the chimney (up to around 25 ppm), and a gradual decrease to values below 1 ppm at a certain distance from the chimney. Since kriging achieves the minimisation of the estimation variance, a by-product is the kriging variance, or its square root ; the kriging standard deviation σ_K. If the kriging error (i.e., the difference between the kriged and true values) is assumed Gaussian, confidence intervals can be deduced from σ_K. We write :

$$R_K - t\sigma_K < R < R_K + t\sigma_K \tag{7}$$

where t is a factor which depends on the level of confidence assigned to R. For a level of confidence of 95%, for example, $t = 1.65$: R has 95% chances of being lower than $R_K + 1.65\ \sigma_K$. An estimation of soil concentration is obtained by combining Equations (4), (6) and (7) :

$$\exp\left(R_K + \ln\left(\mathrm{Cd_s}\right)_\mathrm{T}\right)\cdot\left(\exp(\sigma_K)\right)^{-t} < \mathrm{Cd_s} < \exp\left(R_K + \ln\left(\mathrm{Cd_s}\right)_\mathrm{T}\right)\cdot\left(\exp(\sigma_K)\right)^{t} \tag{8}$$

Since the first term in Equation (8) is the median estimator for $\mathrm{Cd_s}$ (Equation 6), $\exp(\sigma_K)$ can be considered as a multiplicative standard deviation. It is called here an "error factor" for short. The calculated error factor has values up to 1.3 with hollows centred around measurement points. At these points, the remaining uncertainty is due to measurement errors. The statistical soil cadmium concentrations which result from this analysis are combined below with fuzzy uncertainties related to the uptake of cadmium by vegetables and to the absorption of a dose by a human target.

Measured values of cadmium in leeks are plotted in Fig. 8 as a function of measured soil concentrations at the locations where the leeks were grown. As seen in this figure, the small number of measurements (five) hardly warrants a statistical analysis. They nevertheless provide very valuable information as they represent site-specific values of metal uptake by vegetables. This information can be analysed using a fuzzy correlation. Based on the shape displayed by the measured data, it was chosen not to address plant uptake through the use of a "bioconcentration factor" such as appears in several risk analysis tools (for example HESP ; Poels et al., 1990). Such a factor assumes that as soil concentration increases, plant uptake must increase proportionally. But agronomists argue that plant uptake cannot increase indefinitely, and that it will not exceed a certain level. Therefore a log-shaped correlation seemed more appropriate, and consistent with the data depicted in Fig. 8. The following correlation equation is defined :

$$\mathrm{Cd}^*_\mathrm{pl} = Or + (As - Or)\left(1 - \exp\left(-k\,\mathrm{Cd_s}\right)\right) \tag{9}$$

where $\mathrm{Cd_{pl}}^*$ = estimated cadmium concentration in the plant (mg Cd / kg dry plant); $\mathrm{Cd_s}$ = measured cadmium concentration in the soil (mg Cd / kg dry soil); $Or = \mathrm{Cd_{pl}}^*$ at the origin ($\mathrm{Cd_s} = 0$); As = Asymptote ($\mathrm{Cd_{pl}}^*$ at large values of $\mathrm{Cd_s}$); k = parameter which controls the rate of increase. Parameters Or, As and k in Equation (9) are represented by fuzzy numbers, while the $\mathrm{Cd_s}$ values are represented by probability density functions derived from the geostatistical analysis. In Fig. 8, the dashed lines (from Equation 9) represent "likely" boundaries for cadmium uptake by leeks. These boundaries imply that if someone were to go and measure the concentration in leeks cultivated on this site, it is likely that he/she would obtain a value which falls within these boundaries. The full line represents what is considered as a reasonably conservative upper boundary for cadmium concentration in leeks grown on this site. While one could argue regarding the precise positions of these curves, it should be noted that this representation is at least consistent with the measured data, and that it can be easily adjusted to accommodate input from agronomic experts. The fuzzy numbers for parameters *Or*, *As* and *k* are presented in Fig. 9 (along with another parameter discussed below). The limits of these fuzzy numbers are deduced directly from Fig. 8. For example *Or* ; the cadmium content of the plant at the origin ($Cd_s = 0$), is considered likely to be between 0 and 0.3 (these values have a likelihood of 1 in Fig. 9.A) while a value of 0.6 is considered as a possible upper boundary (in Fig. 9.A, the likelihood of values above 0.6 is considered nil).

The objective of the calculation is to estimate a dose of cadmium absorbed by a human target. In fact, doses are calculated along a grid which covers the site in order to examine the spatial distribution of absorbed dose. The dose is calculated from :

$$Dose = \frac{\mathrm{Cd}_{\mathrm{pl}}^{*} \cdot 1000 \cdot Con \cdot DMC}{BW} \tag{10}$$

where Dose = Absorbed dose (μg Cd per day per kg body weight); *Con* = Leek daily consumption (kg leek per day); *DMC* = Leek dry matter content (weight percent); *BW* = Human target body weight (kg). The leek dry matter content (*DMC*) accounts for the fact that daily consumption is provided with respect to wet weight while the cadmium concentration in the leeks (Cd_{pl}^*) is relative to dry weight. The cadmium concentration in the leeks is obtained from the fuzzy correlation equation (Equation 9). As a simplifying hypothesis, it is considered that leeks are representative of vegetables with respect to human exposure to cadmium through vegetable consumption. The daily vegetable consumption and the vegetable dry matter content are selected based on data presented in INERIS (1999). Likely daily vegetable consumption is taken between 100 and 120 g/day, while an upper possible limit is taken as twice the higher value (240 g/day). The vegetable dry matter content (*DMC*) and the body weight are considered as constant : 15% and 70 kg respectively. It is assumed conservatively that the dose absorbed by the target is not influenced by vegetable preparation or assimilation.

An example calculation performed with the hybrid approach is presented for a median value of $Cd_{sM}^* = 7.97$ ppm (from Fig. 7), and its corresponding error factor = 1.15. The probability distribution for Cd_s is obtained by calculating a normal Gaussian distribution for mean = ln(7.97) and standard deviation = ln(1.15), and then taking the exponential of the results to return to the distribution on Cd_s. The hybrid approach is performed according to Figs. 3 and 4 (which represent the general case). Note that as there is only one probabilistic variable involved, and because Equation (10) is very simple (minima and maxima can be identified directly), the calculation can also be performed without Monte Carlo random sampling. Using the values for Cd_{sM}^* and the error factor above, the maximum value of Cd_s for a 95 % confidence level is obtained from Equation (8) for $t = 1.65$: $Cd_{s\ max} = 10.04$ ppm. Likewise, the minimum value of Cd_s is obtained for $t = -1.65$: $Cd_{s\ min} = 6.33$ ppm. These values then serve in Equations (9) and (10) to obtain the *Inf* and *Sup* values of the Dose, for each value of α (level of likelihood), using the relevant *Inf* and *Sup* values of the fuzzy numbers involved in Equation (10). The general hybrid approach, with Monte Carlo sampling, was also performed in order to check that results were identical. The results of this specific calculation are depicted in Fig. 10. Calculations were performed for a large number of points using the same grid as the one used to generate Fig. 7.

The next step consisted in examining the acceptability of the calculated doses. This acceptability is considered here with respect to a maximum reference dose. According to

WHO (1994), the kidney is the main target of cadmium toxicity. In order to maintain cadmium concentrations in the kidney cortex below 50 mg/kg, WHO (1994) recommends that cadmium absorption via food consumption should not exceed 1 μg per day and per kg body weight. The calculated fuzzy doses were compared to this daily reference dose (noted D_o) using the measure of possibility (see Dubois and Prade, 1988) for the proposition : calculated fuzzy dose *F* exceeds reference dose D_o. For a "crisp" reference dose, the measure of possibility is written :

$$\Pi(F > D_o) = \underset{u > D_o}{Sup}\ \mu_F(u) \tag{11}$$

where $\mu_F(u)$ = membership function of *F* for any value *u*; *Sup* = the largest value. Fig. 11 provides a graphical illustration of Π for such a proposition. As long as the fuzzy dose is entirely below the reference dose, the possibility of D_o being exceeded is considered nil (Π = 0 ; Fig. 11a). As the fuzzy dose intersects the reference dose, excess of D_o is considered possible with a possibility measure Π = α (Fig. 11b). Once the reference dose intersects the plateau (Fig. 11c), Π becomes equal to 1. Note that the latter case does not imply that the reference dose will be exceeded with "certainty". Reasonable certainty occurs when another indicator of the validity of the proposition, i.e. the measure of necessity (Dubois and Prade, 1988), becomes equal to 1 :

$$\Pi(F > D_o) = 1 - \underset{u < D_o}{Sup}\ \mu_F(u) \tag{12}$$

Applying the possibility indicator to the calculated fuzzy doses, we obtain the spatial distribution of the possibility of exceeding the reference dose. This spatial distribution is depicted in Fig. 12. Possibilities of 0.45 of exceeding the reference dose are found in the close vicinity of the chimney, and decrease below 0.1 at a certain distance from the chimney.

DISCUSSION AND CONCLUSIONS

The map of Fig. 12 can serve as an aid in the decision-making process related to the management of this industrial site. It would be incumbent, however, upon the health authority, to define the acceptable level of "possibility" of reference dose exceedence. Depending upon the context, values around 0.1-0.2 may seem consistent with a reasonable application of the precautionary principle. From Fig. 12, such values would result in an area on the order of 20 square kilometers being ruled out for vegetable growth. To require a possibility of zero may in many cases be too strict, and result in excessive areas of land being ruled out for certain uses, or in excessive cleanup costs. It is reminded that the fuzzy calculus component of the proposed hybrid approach considers all possible combinations of fuzzy parameter values, and does not transmit through multiplication the uncertainty of these parameter values onto that of the calculation result. As illustrated below, it is more conservative than a purely Monte Carlo calculation, and therefore constraints on acceptance criteria in terms of possibilities need not be as strict as in terms of probabilities.

In order to illustrate the potential consequences of assuming probabilities throughout, the fuzzy numbers in Fig. 9 were converted to probability distribution functions, and hence to PDF's, by adjusting the y-axes so that the integrals beneath the curves equalled unity. Note that this procedure is not the mathematically rigorous manner for establishing a correspondence between fuzzy numbers and probabilities (see Dubois and Prade, 1993), but mimics the *a priori* assumption of attributing probabilities rather than possibilities to the limit values. A high soil cadmium concentration was selected, i.e. a value from the immediate vicinity of the chimney (Cd_{sM}^{*} = 25 ppm, standard deviation = 1.2 ppm). The PDF's and the results of the Monte Carlo simulation (1 000 iterations) are shown in Fig. 13. In Fig. 14, The relative frequencies from the Monte Carlo simulation are compared with the reference dose. Based on this figure, the calculated dose would be considered as "tolerable", even though the soil concentration value is located in the immediate vicinity of the chimney. Thus the entire site would be considered suitable for vegetable growth.

Although the proposed hybrid method may hold some value with respect to a reasonable application of the precautionary principle, and for promoting consistency between uncertainty representation and available data, there are at least two important limitations with respect to its application. Because in the general case (see Figs. 3 and 4), the hybrid approach combines two computationally-intensive methods (Monte Carlo sampling and fuzzy calculus by the method of α-cuts), its application would appear to be limited to calculations involving relatively simple models, such as analytical solutions, rather than numerical models requiring spatial discretisation. There should be scope, however, for such an approach in a risk-assessment framework, as fate and exposure models often involve relatively simple equations.

The second, and probably more serious limitation, relates to the social acceptance of a level of possibility of risk exceedence (see the beginning of this section). Probabilities of risk exceedence have only recently gained some degree of acceptance by regulatory authorities. According to sociologists working in the field of industrial risks (Antoine-Paille, *pers. comm.*), in terms of social communication on risks, probabilities are impossible to "sell" to the general public. In view of these difficulties, it is questionable whether possibilities of risk exceedence will be easily implemented in a regulatory framework. It is felt in any case that definition of possibility cutoff levels would require input from the field of social sciences.

In this paper a hybrid approach was proposed for combining probabilistic and possibilistic representations of model parameter uncertainty. As stated in Guyonnet et al. (1999), if data are available which substantiate a statistical representation of parameter value variability, then such a representation should certainly be preferred. Possibility theory is proposed as an alternative tool to assist in situations where such data are not available, and it is chosen not to force-fit probability distribution functions on data without statistical justification. The proposed hybrid approach presents the double advantage of preserving the strengths of statistical analysis, while providing the flexibility of the fuzzy approach when such an analysis is not substantiated by the data. It is proposed as a means to improve consistency between calculation hypotheses and available information.

As was stated previously, the hybrid approach is only one alternative among others for combining different modes of representation of uncertainty in a same calculation. Further research is needed in order to identify the pros and cons of each alternative. Another interesting issue which will be investigated in the near future, is the inverse problem. In the application example presented above, the hybrid approach was applied in a forward mode to evaluate the possibility of exceeding a tolerable reference dose. The inverse problem is : given a tolerable reference dose, and considering the uncertainty (probabilistic and possibilistic) affecting model parameter values, which residual soil concentration is such that the "possibility" of reference dose exceedence stays below the value fixed by the health authority ? This issue is of particular relevance to risk-based corrective action.

Acknowledgements

We thank Métalleurop France for kindly providing some of the soil data analysed in this paper. We are also grateful to the anonymous reviewers, whose comments helped us to clarify our original manuscript. This study was funded in part by the French Ministry of the Environment, within the GESSOL Programme led by the French National Institute for Agronomic Research (INRA). Funding was also provided by the BRGM's Directorate of Research Activities.

APPENDIX I. REFERENCES

Bardossy, A., Bronstert, A., and Merz, B., (1995). "1-, 2- and 3-dimensional modeling of groundwater movement in the unsaturated soil matrix using a fuzzy approach." *Advances in Water Resources*, 18(4), 237-251.

Bonano, E., Hora, S., Keeney, R., von Winterfeldt, D. (1990). "Elicitation and use of expert judgement in performance assessments for high-level radioactive waste repositories." *Sandia National Laboratories* Report SAND89-1821, Albuquerque, New Mexico.

Cazemier, D. (1999). "Utilisation de l'information incertaine dérivée d'une base de données sols. (Use of uncertain information derived from a soil data-base)." *Ph.D. Thesis of the National School of Agronomy*, Montpellier (France). 170 pp.

Chaney, R., Brow, S., Stuczynski, T., Daniels, W., Henry, C., Li, Y.-M., Siebielec, G, Malik, M., Angle, J., Ryan, J., and Compton, H. (1999). "In-situ remediation and phytoextraction of metals from hazardous contaminated soils." *In* : Innovative clean-up approaches : Investments in technology developments, results & outlooks for the future. Nov.2-4, Bloomingdale, USA. 29 pp.

Chang, A., Page, A., and Warneke, J. (1987). "Long-term sludge application on cadmium and zinc accumulation in Swiss chard and radish." *Journal of Environmental Quality,* 16, 217-221.

Chilès, J.-P., and Delfiner, P. (1999). *Geostatistics: Modeling Spatial Uncertainty*. Wiley, New York.

Cullen, A.C., and Frey, H.C. (1999). "Probabilistic techniques in exposure assessment : a handbook for dealing with variablility and uncertainty in models and inputs." Plenum, New York.

Dijkshoorn, W., Lampe, J., and VanBroekhovern L. (1983). "The effect of soil pH and chemical form of nitrogen fertilizer on heavy metal contents in ryegrass." *Fertilizer Research,* 4, 63-74.

Dou, C., Woldt, W., Bogardi, I., and Dahab, M. (1995). "Steady-state groundwater flow simulation with imprecise parameters." *Water Resources Research,* 31(11), 2709-2719.

Dubois, D., and Prade, H. (1993). "On possibility/probability transformations." *In* : Fuzzy Logic : State of the Art (Lowen, R., Roubens, M. eds.). Kluwer Academic Press, Dordrecht.

Dubois, D., and Prade, H. (1992). "When upper probabilities are possibility measures." *Fuzzy Sets and Systems,* 49, 95-74.

Dubois, D., and Prade, H. (1988). *Possibility theory*. New York Plenum Press, 263 pp.

Freissinet, C., Erlich, M., and Vauclin, M. (1998). "A fuzzy logic-based approach to assess imprecision of soil water contamination modelling." *Soil & Tillage Research,* 47, 1-17.

Guyonnet, D., Côme, B., Perrochet, P., and Parriaux, A. (1999). "Comparing two methods for addressing uncertainty in risk assessments." *Journal of Environmental Engineering*, 125(7), 660-666.

INERIS (1999). "Méthode de calcul des valeurs de constat d'impact dans les sols. (Method for calculating risk-based concentration limits in soils)." *INERIS Unpublished Report*, April 1999, Verneuil-en Halatte, France.

Jopony, M., and Young, S. (1993). "Assessment of lead availability in soils contaminated by mine spoil." *Plant and Soil*, 151, 273-278.

Labieniec, P. Dzombak, D., and Siegrist, R. (1997). "Evaluation of uncertainty in a site-specific risk assessment." *Journal of Environmental Engineering,* 123(3), 234-243.

Lorenz, S., Hamon, R., Holm, P., Domingues, H., Sequeira, E., Christensen, T., and McGrath, S. (1997). "Cadmium and zinc in plants and soil solutions from contaminated soils." *Plant and Soil*, 189, 21-31.

Luttringer, M., and de Cormis, L. (1979). "La pollution par les métaux lourds à Noyelles-Godault et ses environs (Pas de Calais). (Pollution by heavy metals at Noyelles-Godault and surrounding area, Pas de Calais)." *Unpublished report of the National Institute for Agronomic Research (INRA)*, Montfavet (France), 12 pp.

Oreskes, N., Shrader-Frechette, K., Belitz, K. (1994). "Verification, validation, and confirmation of numerical models in earth sciences." *Scienc*e, 263, 641-646.

Poels, C., Gruntz, U., Isnard, P., Riley, D., Spiteller, M., ten Berge, W., Veerkamp, W, Bonyinck, W. (1990). "Hazard assessment of chemical contaminants in soil." ECETOC Report No. 40, ISSN-0773-8072-40. European Chemical Industry Ecology & Toxicology Center, Brussels, Belgium.

Prado, P., Draper, D., Saltelli, S., Pereira, A., Mendes, B., Eguilior, S., Cheal, R., Tarantola, S. (1999). "Gesamac : Conceptual and computational tools to tackle the long-term risk from nuclear waste disposal in the geosphere." *European Commission Report* EUR 19113 EN. Office for Official Publications of the European Communities, Luxemburg.

Shafer, G. (1976). *A Mathematical Theory of Evidence*. Princeton University Press.

Singh, B., Narwal, R., Jeng, A., and Almas, A. (1995). "Crop uptake and extractability of cadmium in soils naturally high in metals at different pH levels." *Commun. Soil Sci. Plant Anal*., 26(13&14), 2123-2142.

Smets, P. and Kennes, R. (1994). "The transferable belief model." *Artificial Intelligence,* 66, 191-234.

Smilde, K., Van Luit, B., and Van Driel, W. (1992). "The extraction by soil and absorption by plants of applied zinc and cadmium." *Plant and Soil*, 143, 233-238.

Tessier, A.P., Campbell, G.C., and Bisson, M. (1979). "Sequential extraction procedure for speciation of particulate trace metals." *Analytical Chemistry,* 51, 844-850.

Vose D. (1996). *Quantitative risk analysis - A guide to Monte-Carlo simulation modelling*. Wiley, New York.

WHO (1994). "Quality directives for drinking water. Volume 1 : recommendations. 2nd Edition." *World Health Organisation*. Geneva, Switzerland. ISBN 92 4 254460 4, 202 pp.

Wonneberger, S., Kistinger, S., Deckert, A. (1995). "Unbiased guess, a concept to cope with fuzzy and random parameters ?" *European Commission Report* EUR 16199 EN. Office for Official Publications of the European Communities, Luxemburg.

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

Zadeh, L. (1978). "Fuzzy sets as a basis for a theory of possibility." *Fuzzy Sets and Systems,* 1, 3-28.

APPENDIX II. NOTATION

The following symbols are used in this paper :

As = Asymptote for Equation (9);
BW = body weight;
Cd_{pl} = measured cadmium concentration in the plant;
Cd_{pl}^* = estimated cadmium concentration in the plant;
Cd_s = measured cadmium concentration in the soil;
Cd_{sM}^* = median estimator of cadmium concentration in the soil;
Con = leek daily consumption;
d = distance to the chimney;
D_o = daily reference dose;
DMC = Leek dry matter content (weight percent);
Inf = smallest value;
$F_1, \ldots, F_m$ = m model parameters each represented by a fuzzy number;
k = constant controlling the rate of increase in Equation (9);
$\ln(Cd_s)$ = logarithm of measured cadmium concentration in the soil;
$\ln(Cd_s)_T$ = value of $\ln(Cd_s)$ predicted by a regression equation (trend);
M = model;
Or = Cd_{pl}^* at the origin in Equation (9);
$P_1, \ldots, P_n$ = n model parameters each represented by a PDF;
PDF = probability density function;
R = residual;
Sup = largest value;
α = value of the membership function μ;
α-cut = all values of parameter X within shaded area in Fig. 2;
$\mu_F(u)$ = membership function of F for any value u;
χ = random number;
σ_K = kriging standard deviation for R;
ω = number of Monte Carlo iterations;
Π = possibility measure.

APPENDIX III. FIGURES

List of Figure legends

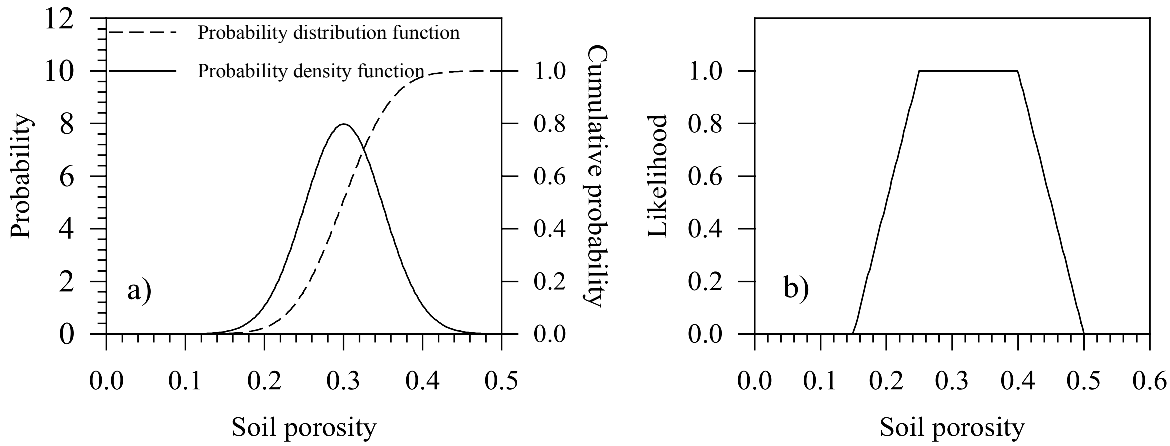

FIG. 1. Illustration of : (a) probability functions, and (b) a fuzzy number

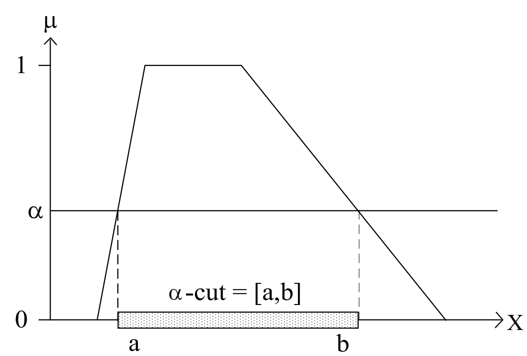

FIG. 2. Illustration of an α-cut; i.e., all values comprised between *a* and *b*

Generate n random numbers (χ) to sample the n PDF's : obtain n values $p_1, \ldots, p_n$.

Apply method of α-cuts : look for *Inf* and *Sup* values of $M(p_1, \ldots, p_n, F_1, \ldots, F_m)$ on the α-cuts. Build fuzzy number of $M(p_1, \ldots, p_n, F_1, \ldots, F_m)$.

Repeat the procedure ω times : obtain ω fuzzy results of $M(P_1, \ldots, P_n, F_1, \ldots, F_m)$.

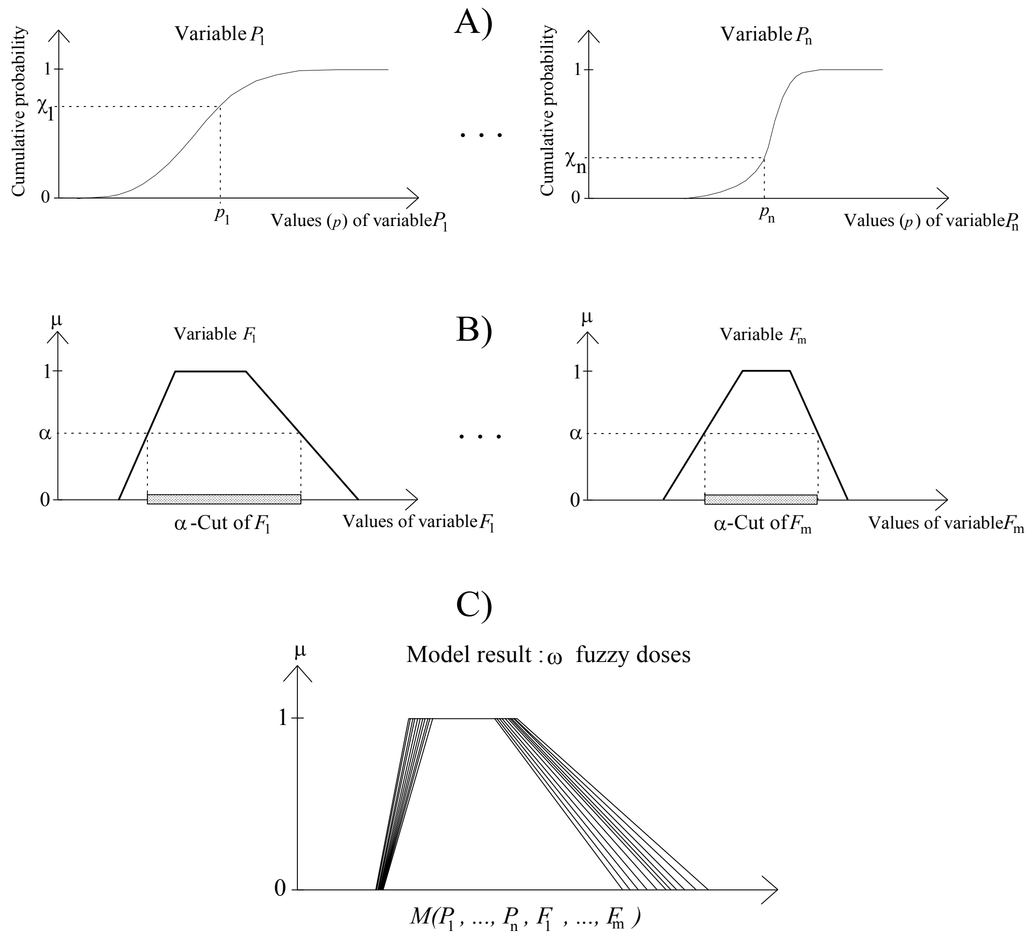

FIG. 3. Schematic illustration of the hybrid approach

Select a value of α and build the cumulative relative frequency diagrams of the *Inf* and *Sup* values of $M(p_1, \ldots, p_n, F_1, \ldots, F_m)$.

Select final *Inf* and *Sup* values of $M(p_1, \ldots, p_n, F_1, \ldots, F_m)$ for 95% confidence level.

Obtain final fuzzy result of $M(p_1, \ldots, p_n, F_1, \ldots, F_m)$ by repeating for each value of α.

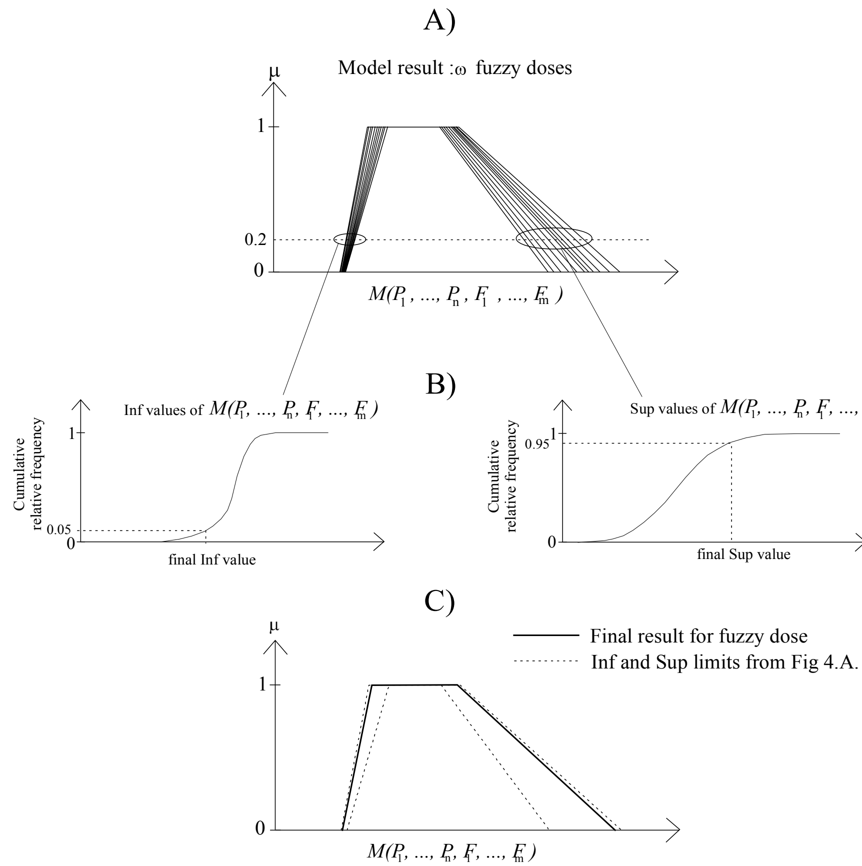

FIG. 4. Selection of the final fuzzy result

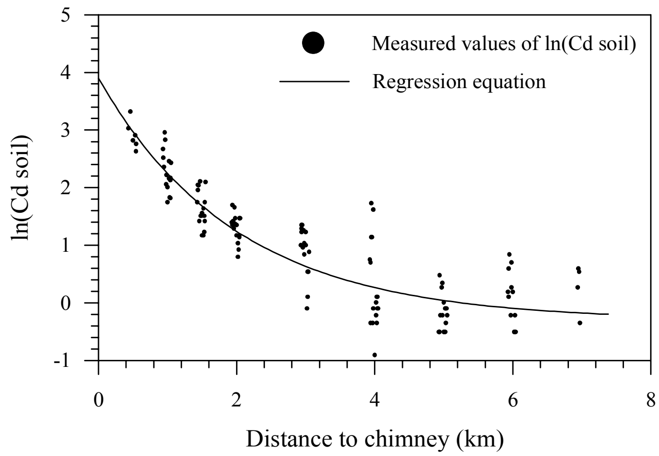

FIG. 5. Correlation between the logarithm of Cd soil concentrations, and distance between the chimney and the sampling point

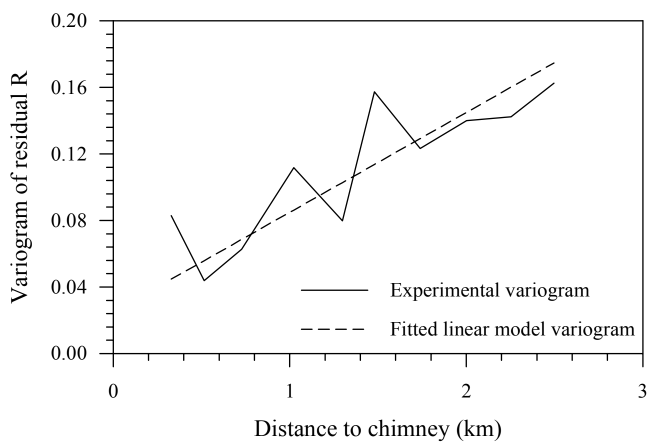

FIG. 6. Variogram of residual *R*. Fit with a linear variogram model

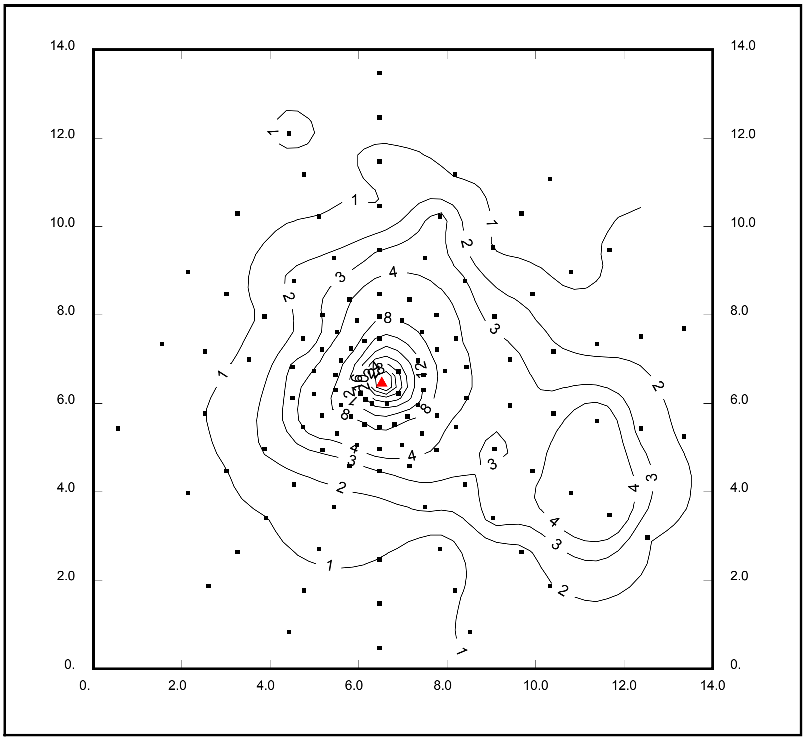

FIG. 7. Interpolated map of soil cadmium concentrations (Cd_s^* ; ppm).
Graduation in km; Triangle = chimney location; Points = measurement points.

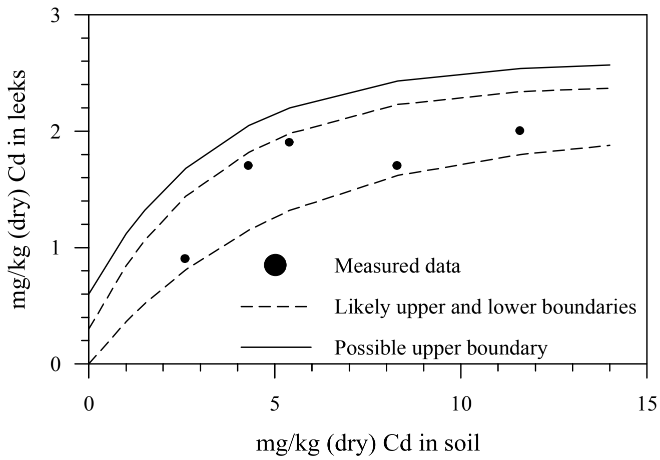

FIG. 8. Cadmium concentrations measured in leeks versus measured soil concentrations, and fuzzy correlation

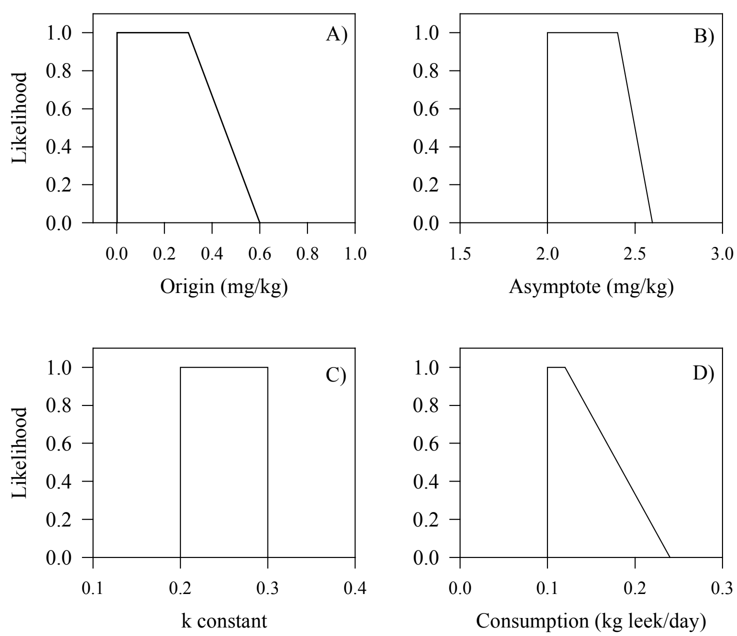

FIG. 9. Fuzzy numbers for several model parameters

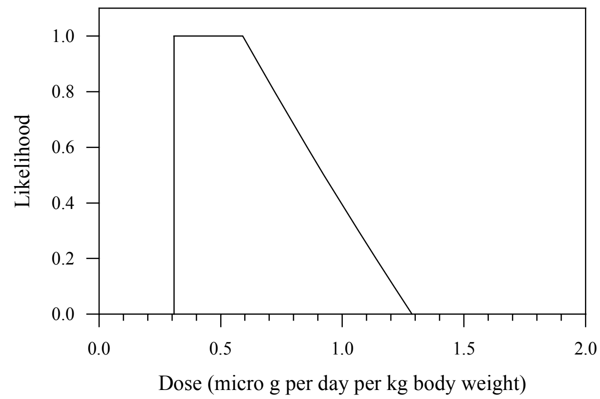

FIG. 10. Calculated dose for Cd_{sM}^* = 7.97 mg/kg, error factor = 1.15 and 95% confidence level on estimated soil cadmium concentration.

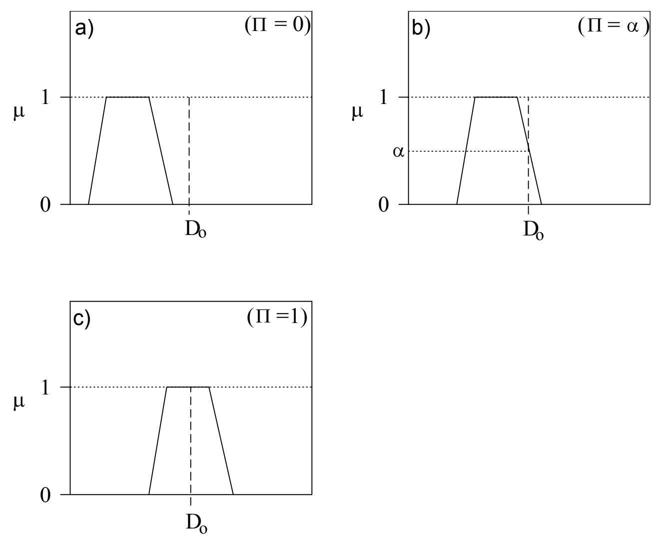

FIG. 11. Comparison between a fuzzy dose and a reference dose (D_o)

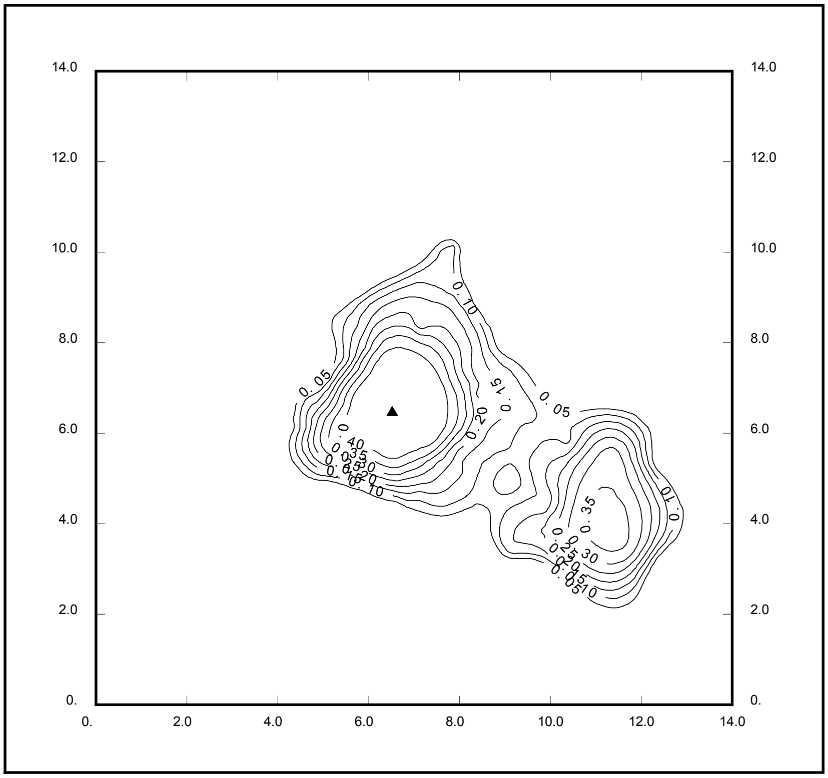

FIG. 12. Map of the possibility that the absorbed dose should exceed the reference dose (1 μg/d kg^{-1}). Graduation in km; Triangle = chimney location.

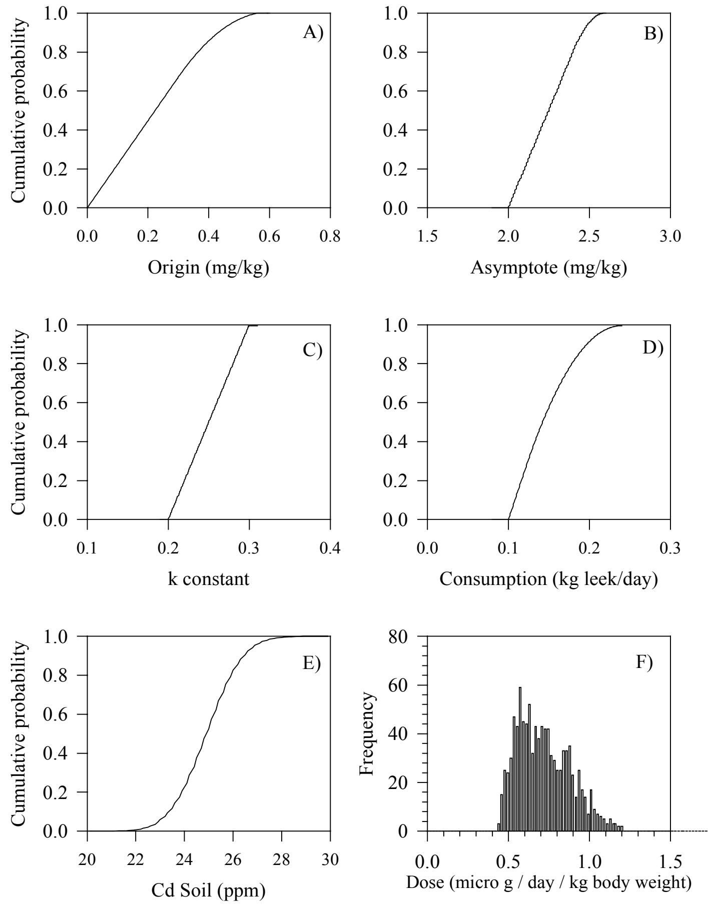

FIG. 13. A) through E) : PDF's based on the shapes of the fuzzy numbers in Fig. 9. F) : result of the Monte Carlo simulation (1000 iterations).

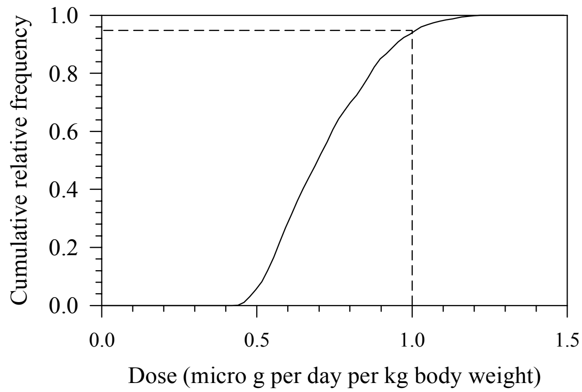

FIG. 14. Comparison between the dose distribution from the Monte Carlo simulation and the reference dose