Theme C: SITE CHARACTERISATION & RISK ASSESSMENT

ASSESSMENT OF CHILD EXPOSURE TO LEAD ON AN IRONWORKS BROWNFIELD: UNCERTAINTY ANALYSIS

Cédric Baudrit[1], Dominique Guyonnet[2], Hafid Baroudi[3], Sébastien Denys[3],
Philippe Begassat[4]

1: IRIT, University Paul Sabatier of Toulouse118 route de Narbonne, 31062 Toulouse,
2: BRGM, 3 av. C. Guillemin, 6009 Orléans Cédex 2,
3: INERIS, BP 2, 60550 Verneuil-en-Halatte
4: ADEME, 2 square La Fayette, BP 406, 49004 Angers Cedex 01.

Keywords : risk assessment, lead, uncertainty, PDF, Dempster-Shafer, dependencies.

INTRODUCTION

Uncertainty regarding parameters involved in health risk assessment models is generally addressed within a purely probabilistic framework: it is assumed that knowledge regarding model parameters is of a purely random nature (expressing variability). Such knowledge is represented by single probability distributions (PDFs) and the uncertainty is typically propagated through the risk model using the Monte-Carlo technique. Yet in real cases of site investigations and risk assessments, information regarding model parameters is often vague or incomplete (due to time and financial constraints). Such information does not warrant the use of single PDFs that are truly representative of available information. For example, in their analysis of the uncertainty in the propagation of chlorinated organic pollutants in groundwater, McNab et al. (1999) define 12 probability distributions for their model parameters, 9 of which are qualified as "postulated". As shown in particular by Ferson & Ginzburg (1996) assuming random variability when faced with partial incomplete/imprecise information (partial ignorance) seriously biases the outcome of the analysis by reducing the range of possible outcomes. These authors suggest that distinct methods are needed to adequately represent and propagate random variability (often referred to as "objective uncertainty") and imprecision (often referred to as "subjective uncertainty").

In practice, while information regarding variability is best conveyed using probability distributions, information regarding imprecision is more faithfully conveyed using families of probability distributions encoded either by probability boxes (upper and lower cumulative distribution functions (Ferson et al., 2003; Ferson & Ginzburg, 1996) or by possibility distributions (also called fuzzy numbers) (Dubois & Prade, 1988, 1992, 2000) or yet by belief functions of Shafer (1976). Previous authors have shown how these two types of uncertainty can be represented (Baudrit, et al., 2004, Ferson et al., 2003a, 2003b) and propagated in the estimation of risk (Baudrit et al., 2003, 2004, Guyonnet et al., 2003). An issue that has not been addressed by many researchers (Ferson et al., 2004) is the influence of possible dependencies between parameters on the results of the risk assessment. Indeed, the presence of imprecision in the knowledge of model parameters potentially generates two levels of dependencies. This aspect is examined here using a real case of soil contamination by lead on an ironworks brownfield

CASE DESCRIPTION

Sources of contamination leading to potential exposure on the ironworks brownfield are:

- Waste water loaded with lead stored in a decanting tank which knew several overflows during periods of strong precipitations,
- The deposition of lead-rich chimney dust at the surface of the soil during site operation,
- The deposition of lead-rich dust at the surface of the soil during the demolition phase of the ironworks factory where materials soiled by lead were burned on site.

Site investigation revealed the presence of lead in the superficial soil at levels on the order of tens of grams per kg dry soil. A risk-based cleanup objective of 300 mg/kg was established by a consulting company, taking into account the most significant exposure pathway and the most sensitive target (direct soil ingestion by children). The risk calculation was performed without considering parameter uncertainty. A reanalysis of this risk assessment is presented below that attempts to take into account such uncertainty. The emphasis is on the influence of possible dependencies between calculation parameters.

The mathematical model that was used to calculated a dose (*D_{lead}*) absorbed by a child living on the site and exposed via soil ingestion, is (USEPA, 1989):

$$D_{lead} = \frac{\left(C_{soil} \times Ir_{soil} \times \left(Fi_{indoor} + Fi_{outdoor}\right) \times \left(Efsoil_{indoor} + Efsoil_{outdoor}\right) \times Ed\right)}{Bw \times At \times 10^{6}} \qquad (1)$$

Where :
D_{lead} = daily lead dose absorbed via soil ingestion (mg Pb/kg body weight per day),
C_{soil} = total lead concentration in soil (mg/kg dry soil),
Ir_{soil} = ingestion rate (mg soil per day),
F_i = contaminated soil ingestion fraction (unitless),
Ef_{soil} = exposure frequency (days per year),
Ed = exposure duration (years),
Bw = body weight (kg),
At = Average time (period over which exposure is averaged ; days).

The World Health Organization suggests a maximum acceptable daily dose for lead of 3.5 μg/[kg.day]. This threshold should therefore be respected following cleanup at a residual concentration of 300 mg/kg. Typically, the model parameters are tainted by objective and subjective uncertainty. In the following section, we propose mathematical representations according to the nature of the available information (variability-imprecision) concerning each model parameter. Next we process the uncertainty on the estimation of the absorbed lead dose *D_{lead}* using different methods for comparison purposes.

INFORMATION REPRESENTATION

With respect to the different parameters that appear in Equation (1), body weight (Bw) was chosen to be represented by a single probability distribution (PDF), selected based on available statistical data. This parameter is represented by a normal PDF of average 17.4 kg and standard deviation 2.57 kg with cut-offs at 12.9 kg and 21.5 kg.

On the other hand, due to the very partial character of available information regarding several other parameters (soil concentration *C_{soil}*, ingestion rate *Ir_{soil}* and contaminated soil ingestion fraction *Fi*), it was preferred to represent these parameters by possibility distributions (fuzzy numbers) established based on expert judgement. It is reminded that such distributions represent families of probability distributions (see Guyonnet et al., 2005; this conference for an illustration of the correspondence between possibility distributions and families of probability distributions).

Considering it unrealistic to assume that the cleanup objective will be met precisely at all locations on the site, a triangular possibility distribution was defined with core (value considered most likely) = 300 mg/kg and support (interval outside which values are judged unrealistic) = [40-500] mg/kg (Figure 1).

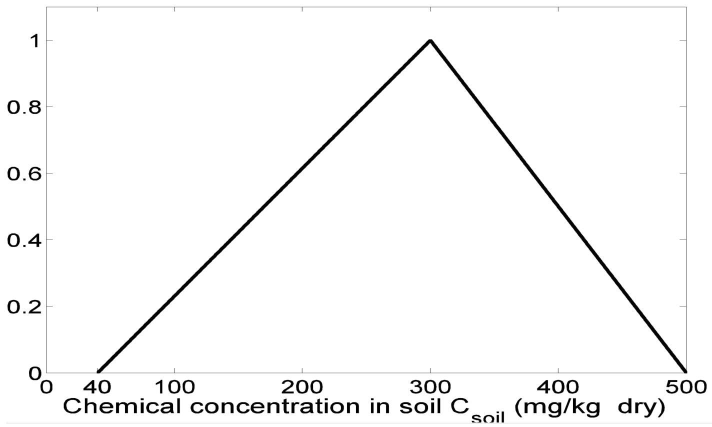

Figure 1 - Triangular possibility distribution representing soil concentration after cleanup

Note that for the purely probabilistic calculation performed below, this shape of distribution will be assumed to represent a probability density function. It should be also noted that this probability density function is just one representative of the family of functions encoded by the possibility distribution of Figure 1. It can be shown that this triangular possibility distribution defines a family of probabilities which contains all unimodal probabilities of mode 300 mg/kg limited by 40 mg/kg and 500 mg/kg (Dubois et al., 2004).

With respect to Ingestion rate Ir_{soil} (mg soil/day), experts suggested representing this parameter using a possibility distribution (Figure 2).

Experts consider that child ingestion rate is most likely to lie between 100 and 150 mg of soil per day but they do not exclude values as low as 40 and as high as 300 mg per day. The observed convexity of the possibility distribution conveys the idea that values outside the core [100-150] are judged less likely than if linear branches were used. While it can be argued that the choice of this shape is subjective, the consequences of such subjectivity on the outcome of the analysis is far less than if single PDFs were selected.

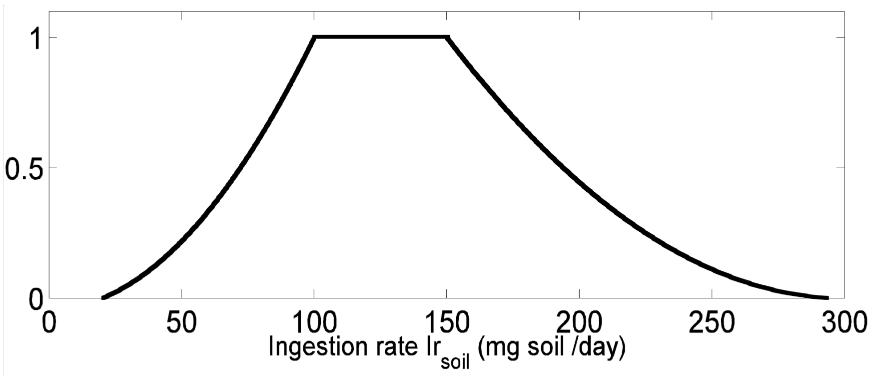

Figure 2 - Possibility distribution representing ingestion rate

The contaminated soil ingestion fraction (*Fi*) decomposes into a fraction ingested indoor Fi_{indoor} and a fraction ingested outdoor $Fi_{outdoor}$. The outdoor ingestion fraction is taken as unity, while the indoor ingestion fraction is defined by Figure 3 based on expert judgement.

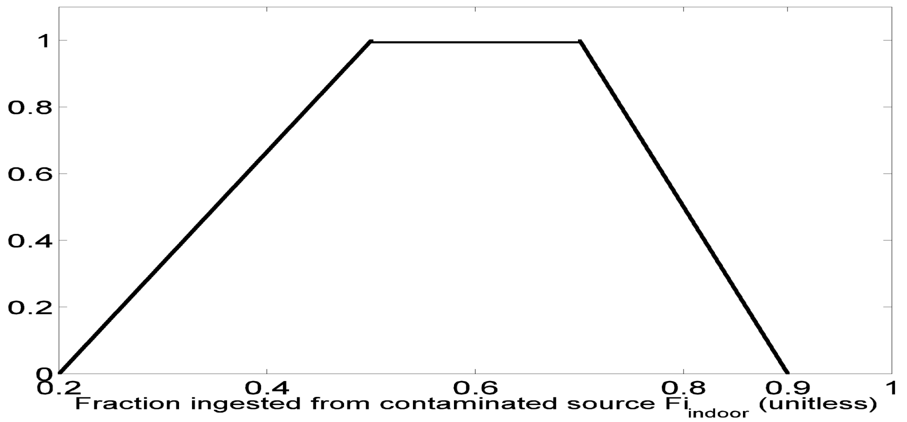

Figure 3 - Trapezoidal possibility distribution representing fraction ingested indoor

Child exposure duration is fixed at 6 years. Exposure frequency is decomposed into an indoor exposure $Efsoil_{indoor}$ equal to 16 hours per day and an outdoor exposure $Efsoil_{outdoor}$ equal to 2 hours per day. Exposure is averaged over 6 years (*At* = 2190 days). the period over which exposure is averaged is taken as the exposure duration, as carcinogenic effects for lead are not proven

UNCERTAINTY PROPAGATION

The first method of propagation is the classical interval analysis. This approach defines a “worst case”, but provides no discrimination within the obtained interval. Should the worst case be lower than the prescribed threshold, the analysis is sufficient. But if this is not the case, a graduation of evidence within the interval is necessary.

For the interval analysis, all model parameters are represented by [min - max] intervals (defined by the supports of the distributions presented previously; Table 1) and we calculate the set of possible values of lead dose D_{lead} related to soil ingestion. This approach provides a bracketing of lead dose D_{lead} but does not provide information regarding the extent to which the criterion D_{lead} < 3.5 µg/[kg day] is satisfied. Table 1 presents the maximum and the minimum values of lead dose D_{lead}. According to this interval analysis, the “worst case” D_{lead} exceeds the limit recommended by WHO by a factor 5.

Parameter	Minimum	Maximum
C_{soil} (mg/kg dry)	40	500
Ir_{soil} (mg soil/day)	20	300
Fi_{indoor}(unitless)	1	1
$Fi_{outdoor}$(unitless)	0.2	0.9
Ed (years)	6	6
$Efsoil_{indoor}$(days/year)	2/24 x 365	2/24 x 365
$Efsoil_{outdoor}$(days/year)	16/24 x 365	16/24 x 365
Bw (kg)	12.9	21.5
At (days)	6 x 365	6 x 365
Dlead (µg/[kg.day])	0.0335	16.6

Table 1 - Summary of interval calculation regarding lead dose

The second method applied is the purely probabilistic method. Stochastic uncertainty is propagated through the mathematical model using the Monte-Carlo method with a hypothesis of independency between model parameters. Single distributions are used to describe the uncertain parameters, with the corresponding density functions based on the same shapes as the possibility distributions described in the previous section. Note that this approach provides “precise” results with respect to the probability of exceeding the threshold. But this precision can be judged arbitrary in (Eq.1) since that the use of single probability distributions to represent imprecisely-known model parameters is not justified by available information.

The result in Figure 4 shows the probability that lead dose D_{lead} should exceed the WHO threshold after cleanup. This probability is of 27% implying a probability of 73% that D_{lead} should be lower than the threshold. Such a level of probability might seem close to sufficient in terms of risk. However, as will be shown below, if the imprecise nature of available information is taken into account, the likelihood of exceeding the threshold is found to be far more significant than Figure 4 seems to indicate.

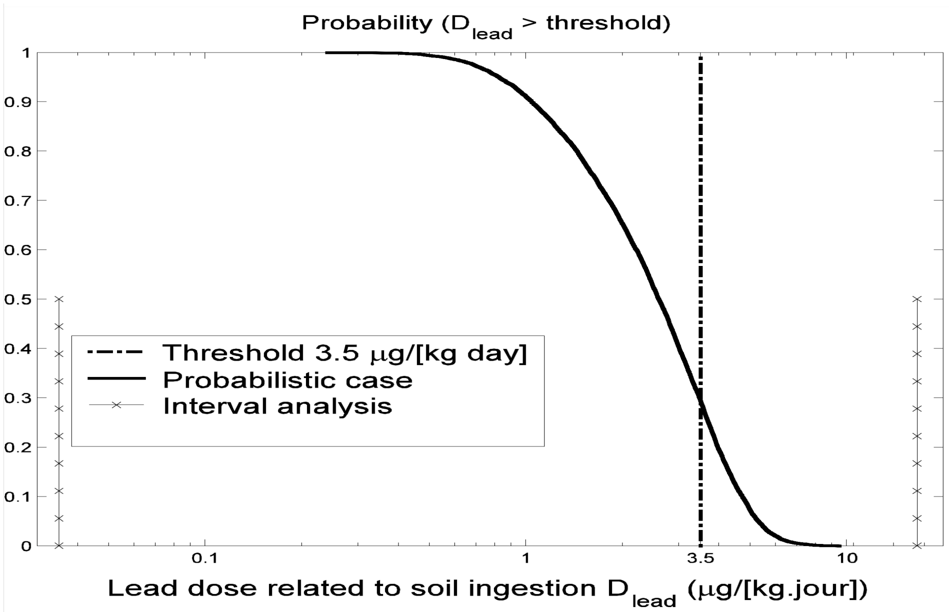

Figure 4 – Uncertainty analysis in a purely probabilistic framework

The methods described below acknowledge the imprecise nature of available information and attempt to jointly propagate variability (relative to body weight) and imprecision in the estimation of D_{lead}. Four methods are implemented :

- The "*Hybrid*" method (see Guyonnet et al., 2005; this conference) that combines Monte-Carlo random sampling with fuzzy calculus (interval calculus performed on intervals with different levels of likelihood). This method assumes strong dependence between information sources pertaining to possibilistic variables C_{soil}, Ir_{soil} and Fi, i.e. on the choice of confidence levels. It assumes independence between stochastic uncertainty pertaining to Bw and the group of epistemic uncertainties pertaining to C_{soil}, Ir_{soil} and Fi.

- An "*Independence Random Sets*" approach (Baudrit et al., 2003) that uses the belief functions of Dempster-Shafer (Shafer, 1976). Belief functions can encode both probability distributions and possibility distributions. Unlike the "*Hybrid*" approach, this method assumes independence between information sources pertaining to possibilistic variables C_{soil}, Ir_{soil} and Fi. As shown below, this assumption implies that we obtain a smaller range of uncertainty related to D_{lead} than with the "*Hybrid*" method. This method is a conservative counterpart of the calculus on probabilistic variables under hypothesis of independence (Couso et al., 2000).

- A "*Conservative Random Sets*" approach (Baudrit et al., 2003) that also uses the belief functions of Dempster-Shafer and solves linear optimization problems. The idea behind the approach is to compute upper and lower probabilities without assuming knowledge on dependencies between stochastic and epistemic uncertainty sources. Hence, dependencies between model parameters are not the same for the different values of lead dose D_{lead}. For each value of lead dose, we estimate, with linear optimization, dependencies between information sources pertaining to variables, which maximise the upper probability and minimise the lower probability.

- A "*Dependency Bounds Convolution*" approach that uses p-boxes (pairs of upper & lower cumulative distribution functions) of Williamson et al., 1990 extended by Ferson et al., 2003, 2004. This method computes extreme upper and lower cumulative distribution functions based on the results of probabilistic models no matter what correlations or statistical dependencies exist between the variables. It uses the theory of copulas (see Nelson, 1999). A copula C is a mapping C:[0,1][0,1]->[0,1] which allows to define the joint probability of two random variables. So Ferson et al., (2003, 2004) combine the first two variables and then combine the bound result with other variables and so on. Although this process provides a bracketing of the ill-known joint probability, it provides less precise results than the "*Conservative Random Sets"* approach because it models "impossible" dependencies among random variables (see Baudrit et al. 2005). Hence the results may be considered as "exaggeratedly conservative".

Figure 5 compares the results of the purely probabilistic approach to the "*Hybrid*" approach when only the imprecision due to C_{soil} is taken into account (according to Figure 1). Results of the "*Hybrid*" approach are presented as an upper bound on probability (Plausibility; Shafer, 1976) and a lower bound on probability (Belief). The gap between these two distributions reflects the imprecision on C_{soil}. The gap is also influenced by the propagation method and the hypotheses on parameter dependence mentioned above. According to Figure 5 there is a 50% level of plausibility of exceeding the threshold (or a 50% level of belief of being lower than the threshold). If the upper probability of exceeding the threshold is selected as the criterion (conservative), then we see that considering imprecision regarding input parameters leads to a clearer rejection of the proposition "D_{lead} < 3.5 μg/[kg day]".

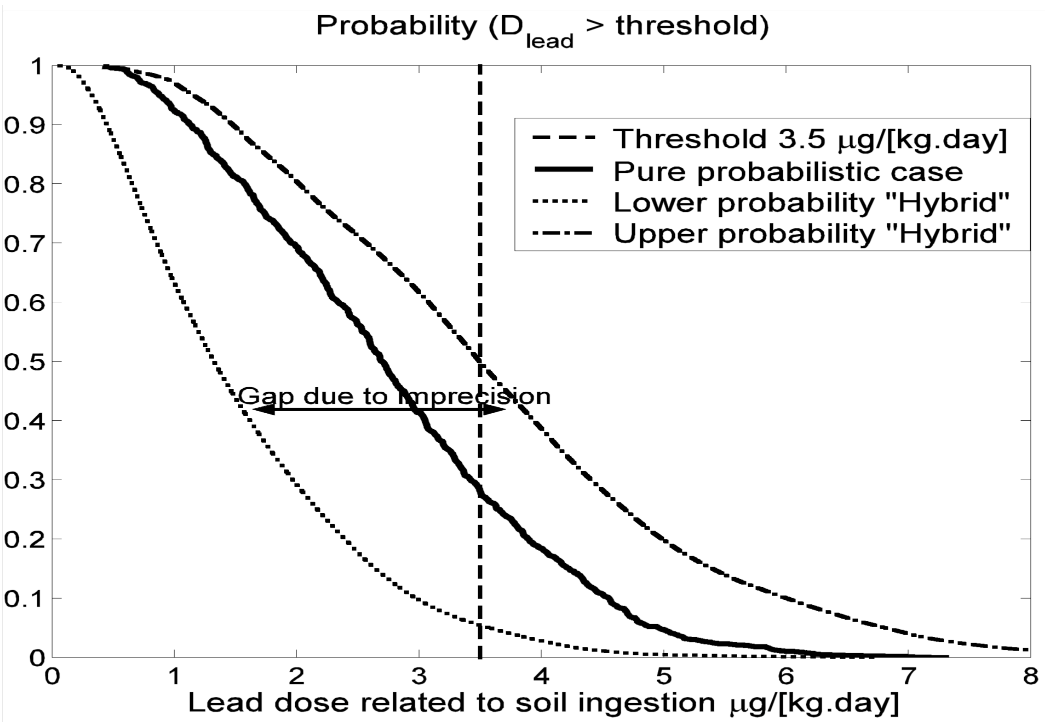

Figure 5 – Comparison between the purely probabilistic method and the "Hybrid" method including imprecision on C_{soil} only

Taking now into account imprecision on ingestion rate Ir_{soil} and on contaminated soil ingestion fraction (*Fi*) (Figures 2 and 3), application of the "*Hybrid*" method yields a much larger gap between the upper and lower probability distributions (Figure 6). According to Figure 6, there is now a very large level of plausibility (90%) of exceeding the threshold.

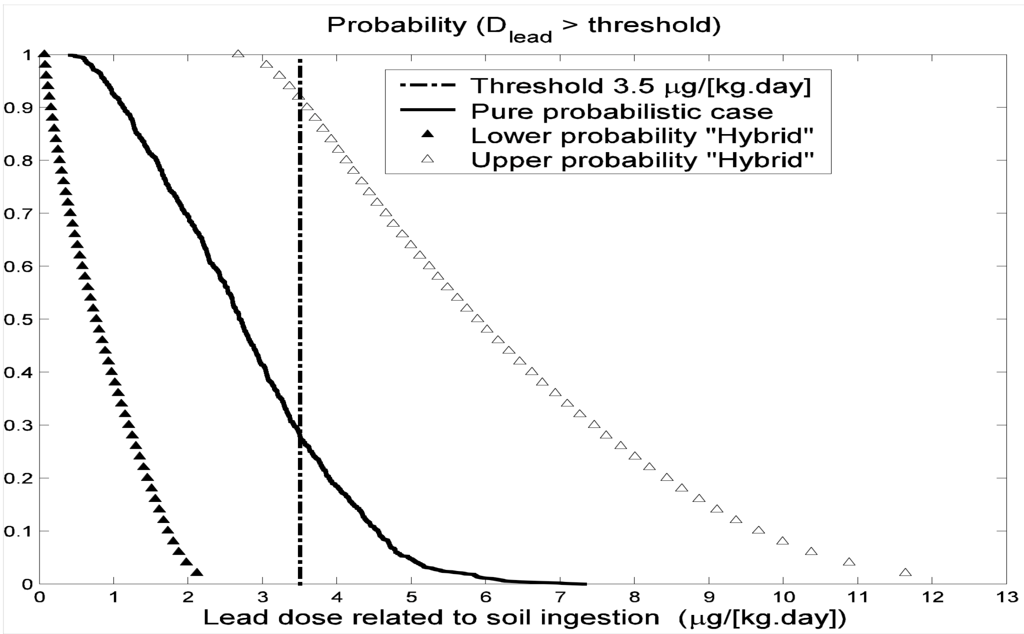

Figure 6 – Comparison between the purely probabilistic method and the “Hybrid” method including imprecision on C_{soil}, Ir_{soil} and Fi.

In the remaining section, we examine the influence of the propagation method on the outcome of the analysis. Each propagation method has its own hypotheses regarding (in)dependence between model parameters. Figure 7 presents the upper and lower cumulative distribution functions according to each propagation method compared to the purely probabilistic case. It indicates that the "*Independence Random Sets*" approach yields a range of uncertainty that is slightly smaller than the one obtained by "*Hybrid*" approach (there is a 23% decrease of the total uncertainty compared to the ”*Hybrid*" approach). We also see that if we make no assumption regarding dependencies between model parameters (“*Dependency Bounds Convolution*") we obtain results that are more conservative than with the "*Conservative Random Sets*" approach. This is due to the fact that the "*Dependency Bounds Convolution*" approach combines probabilities in an iterative fashion two by two and so considers impossible dependencies between random variables.

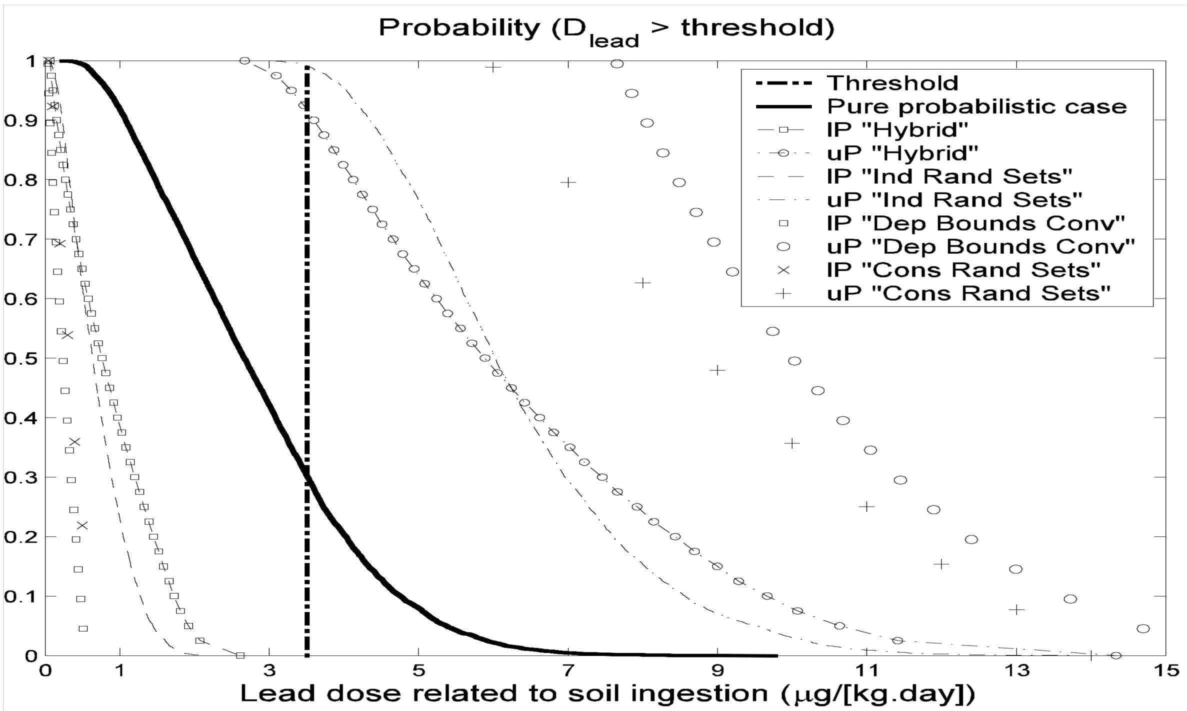

Figure 7 - Upper & lower probabilities of exceeding an arbitrary threshold of absorbed lead dose for different propagation methods. uP = Upper probability, lP = Lower probability

CONCLUSION

This paper addressed the issue of parameter uncertainty in risk assessments with a particular focus on the issue of (in)dependence between model parameters. It was shown that accounting for imprecision in the input information has a large influence on the outcome of the analysis compared to the case where random variability is assumed regardless of imprecise knowledge. Also, different hypotheses regarding dependencies between model parameters have a large influence on the range of possible outcomes as expressed by the gap between the upper and lower probability distributions.

While it might be argued that using single (and subjective) probability distributions is preferable if one wishes to obtain a "precise" answer, it is the opinion of the authors of this paper that consistency with available information should be preferred to "disguising" incomplete information in the form of random variability. Propagating imprecision throughout the analysis may underline the need to go back to acquiring additional information. For example, considering the large gap between the upper and lower bounds on probability in Figure 6, it is clear that further studies on ingestion rates would be needed in order to reduce the subjective uncertainty regarding this parameter (Figure 2). Should single subjective probability distributions be chosen instead, the "precise" probability distribution resulting from the analysis would seem quite sufficient. While in risk analysis there can hardly be any claim to calculation "validity", consistency with available information may be an achievable goal.

ACKNOWLEDGEMENTS

This work is part of Cédric Baudrit's Ph.D. performed under the supervision of Didier Dubois at the University Paul Sabatier of Toulouse, with financial support from BRGM, INERIS and IRSN. Participation of BRGM benefited from a grant by ADEME within the IREA project.

REFERENCES

Baudrit, C., Dubois, D. (2005) – Comparing methods for joint objective and subjective uncertainty propagation with an example in a risk assessment. *Accepted to 4th International Symposium on Imprecise Probabilities and Their Applications*, Pittsburgh, Pennsylvania.

Baudrit, C., Dubois, D., Guyonnet, D., Fargier, H. (2004) - Joint Treatment of imprecision and Randomness in Uncertainty Propagation. *Proc. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems, Perugia* , pp.873-880.

Baudrit, C., Dubois, D., Fargier, H. (2004) - Practical representation of incomplete probabilistic information. *Advances in Soft Computing:Soft Methods of Probability and Statistics conference, Oviedo*}, Springer, 149-156.

Baudrit, C., Dubois, D., Fargier, H. (2003) Propagation of Uncertainty involving Imprecision and Randomness. *Proc.3[e] European Conference on Fuzzy Logic and Technology, Zittau (Germany)*, 653-658.

Couso, I., Moral, S., Walley, P. (2000) - A survey of concepts of independence for imprecise probabilities. *Risk Decision and Policy,* 165-181.

Dubois, D., Mauris, G., Foulloy, L. Prade, H. (2004) - Probability-possibility transformations, triangular fuzzy sets and probabilistic inequalities. *Reliable Computing*, 10, 273-297.

Dubois, D., Nguyen, H. T., Prade, H. (2000) - Possibility theory, probability and fuzzy sets: misunderstandings, bridges and gaps. *Fundamentals of Fuzzy Sets*, Dubois,D. Prade,H., Eds: Kluwer , Boston, Mass, 343-438 .

Dubois, D., Prade, H. (1992) - When upper probabilities are possibility measures. *Fuzzy Sets and Systems*}, 49, 65-74.

Dubois, D., Prade, H. (1988) - Possibility Theory: An Approach to Computerized Processing of Uncertainty, *Plenum Press, New York*..

EPA, U.S. (1989) - Risk Assessment Guidance for Superfund. *Volume I: Human Health Evaluation Manual (Part A). (EPA/540/1-89/002). Office of Emergency and Remedial Response. U.S EPA., Washington D.C.*}

Ferson, S., Ginzburg, L., Akcakaya, R. Whereof one cannot speak: when input distributions are unknown. To appear in *Risk Analysis*.

Ferson, S., Troy Tucker, W. (2003) - Probability Bounds Analysis in Environmental Risk Assessments, *Applied Biomathematics Report*, 2003. http://www.ramas.com/pbawhite.pdf

Ferson, S., Berleant, D., Regan, H.M. (2004) - Equivalent of methods for uncertainty propagation of real-valued random variables. *International Journal of Approximate Reasoning*, 36, 1-30.

Ferson, S., L.Ginzburg, V.Kreinovich, D.M.Myers, K.Sentz. (2003) – Construction Probability Boxes and Dempster-Shafer structures. *Sandia National Laboratories, Technical report SANDD2002-4015*. Available at www.sandia.gov/epistemic/Reports/SAND2002 4015.pdf.

Ferson, S., Ginzburg, L.R. (1996) - Different methods are needed to propagate ignorance and variability. *Reliability Engineering and Systems Safety*, 54, 133-144.

Guyonnet, D., Ménard, Y., Baudrit, C., Dubois, D., Marot, F., (2005) - Joint propagation of variability and imprecision in risk assessments: the HyRisk learnware. ConSoiL2005, 3-7 October 2005, Bordeaux, France.

McNab, W., Rice, D., Bear, J., Ragaini, R., Tuckfield, C., Oldenburg, C. (1999) – Historical case analysis of chlorinated volatile organic compound plumes. *Rapport du Lawrence Livermore National Laboratory, UCRL-AR-133361, Californie, USA*.}.

Nelson, R.B. (1999) - An Introduction to Copulas. *Lecture* Notes in Statistics, Springer Verlag, New York, v. 139.

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

Williamson, R.C., Downs, T. (1990) - Probabilistic arithmetic I: Numerical methods for calculating convolutions and dependency bounds. *International Journal of Approximate Reasoning*, 4, 8-158.