MICA
Minerals Intelligence Capacity Analysis

FACTSHEET

Parameter uncertainty in mineral intelligence analyses

A synthetis of different methods for representing parameter uncertainty in mineral intelligence analyses, with an emphasis on the level of information that is available concerning the parameters.

Scope and description

Uncertainty is an unavoidable aspect of mineral intelligence and the need to address uncertainty can appear in a wide variety of situations. For example when addressing the degree of precision of measurements or when performing modeling calculations based on uncertain parameters. There are many sources of uncertainty; for example data temporal and/or spatial variability, measurement imprecision, model uncertainty, etc. In this factsheet we examine uncertainty regarding data and parameters, with an emphasis on two fundamentally different aspects of uncertainty: stochastic versus epistemic uncertainty (Ferson & Ginzburg, 1996). Stochastic uncertainty, also called "objective uncertainty" arises from the random nature of natural processes. Epistemic uncertainty, also called "subjective uncertainty", arises from the partial character of our knowledge of the natural world. The confusion between these two types of uncertainty is one of the most common pitfalls in uncertainty analysis (Ferson, 1996). The basic message that is underlined herein is that different tools are needed to represent different types of information regarding data uncertainy. If information is "rich" (in the sense of quantity and quality), then standard probability theory should apply. On the other hand if information is poor (scarce or imprecise data, expert judgment, …), then alternative information theories, that were specially developed to handle such information, may be preferred. The underlying objective is to handle information in a manner which is consistent with the nature of that information (Dubois & Guyonnet, 2011). These notions are illustrated below with reference to information regarding Neodymium concentrations in earphones, from end-of-life small audio applications (WEEE; waste electrical and electronic equipment). In Section I, the illustration below takes the reader gradually from "poor" to "rich" information, while Section II relates to the propagation of uncertainty in analyses that are relevant to mineral intelligence capacity.

I Uncertainty representation

I.1 Expert judgment regarding Nd concentrations in earphones

Figure 1 illustrates the use of a simple min-max interval to convey the opinion of an "expert", according to which Neodymium in earphones from small audio applications should lie somewhere between 20 and 140 mg per earphone. The min-max interval is represented in Figure 1 as a "possibility distribution" (or "fuzzy number"; Zadeh, 1978; Dubois & Prade, 1988) which assigns a "membership function" to values within the interval. The min-max interval yields the simplest sort of possibility distribution; i.e. where values of possibility within the interval equal 1 (these values are considered possible) while values outside the interval equal 0 (these are considered not possible).

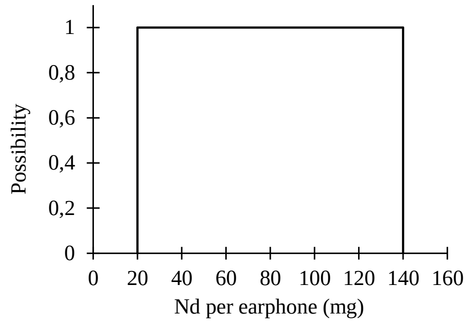

Figure 1 – Interval representing expert judgment regarding possible values of Nd concentrations in earphones

In terms of probabilities, Figure 1 tells us that, according to this expert, there is a 100% probability that Nd concentrations in earphones are lower than 140 mg, and a 0% probability that they are lower than 20 mg. This yields the "family" of cumulative probability distributions depicted in Figure 2.

For values x located within the [20, 140 mg] interval, the probability P(X<x) lies somewhere between 0 and 1: there is no information available to justify more precise values.

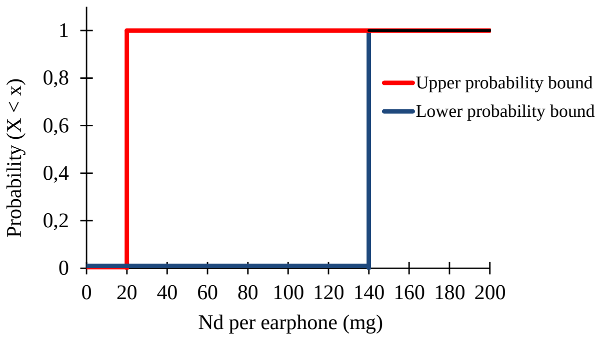

Figure 2 – Upper and lower bounds of the family of cumulative probability distributions corresponding to the information in Figure 1

It could be argued, in a Bayesian framework and applying a "principle of maximum entropy" (Jaynes, 1988), that in the absence of any additional information the uniform probability distribution should be selected to represent uncertainty regarding such a parameter (Figure 3). However, as pointed out by previous researchers, this approach is analogous to selecting just one among all the eligible distributions that respect the information provided by the expert, a few representatives of which are represented in Figure 4.

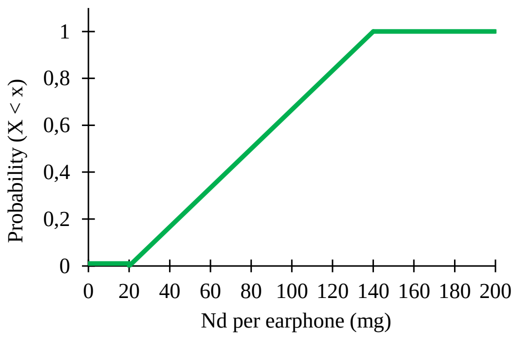

Figure 3 – Hypothesis of a uniform probability distribution over the [min, max] interval

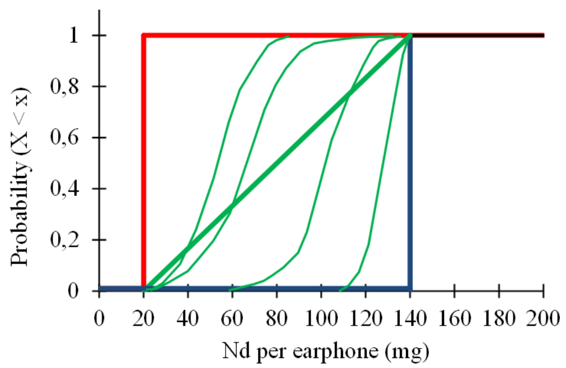

Figure 4 – Some members (in green) of the family of probability distributions delimited by the bounds defined in Figure 2

In many situations an expert is not only able to provide an interval within which he/she feels confident that the values of a parameter should lie, but may also be able to express "preferences" within this interval. The possibility distribution in Figure 5 conveys the following information:

- The expert believes that values of Nd concentrations in earphones lie between 20 and 140 mg/earphone (this interval is termed the "support" of the distribution);
- The expert considers that the most "likely" value is 80 mg/earphone (this constitutes the "core" of the distribution).

It should be noted that the core of the distribution is not necessarily a single value (in which case the distribution is triangular) but can also be an interval of values (in which case the distribution is trapezoidal). Values of possibility assigned to the core of the distribution equal 1, which conveys a notion of "lack of surprise" that actual parameter values should fall within the core.

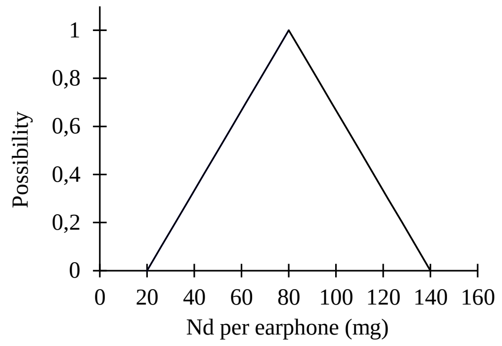

Figure 5 – Possibility distribution over support [20, 140 mg], with a preference expressed for value 80 mg (core)

As for the simple min-max interval, the information in Figure 5 can be conveyed in the form of a family of cumulative probability distributions (Figure 6).

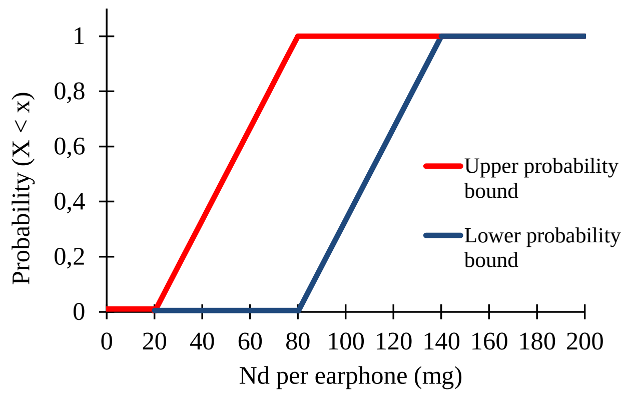

Figure 6 – Upper and lower bounds of the family of cumulative probability distributions corresponding to the information in Figure 5

For example, according to Figure 6, the probability that Nd/earphone is lower than 120 mg, is between 67% and 100% (Figure 7). We will see further below a means to define a single indicator of the evidence of lying below a certain threshold.

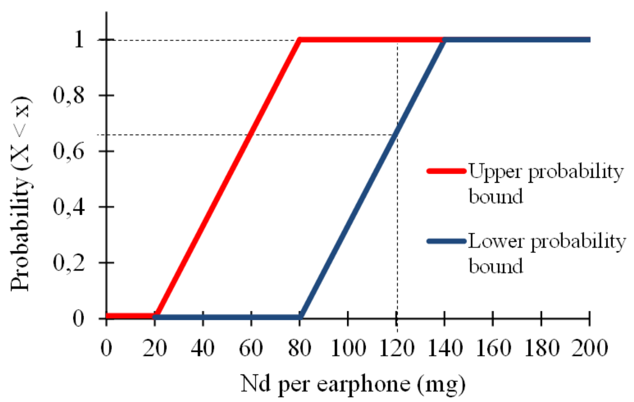

Figure 7 – Probability that Nd content of earphones is lower than 120 mg/earphone

I.2 Scarce data regarding Nd concentrations in earphones

Table 1 shows results of Nd contents in earphones reported by Westphal and Kuchta (2013). The data can ne reported in the form of a cumulative frequency plot (Figure 8)

Table 1 – Nd content of earphones (data from Westphal and Kuchta, 2013)

Unit	mg Nd
Aiwa	75,52
Apple	129,08
Nokia	50,33
Samsung	92,66
Sony Erikson	70,84
Sony I	123,92
Sony II	38,88

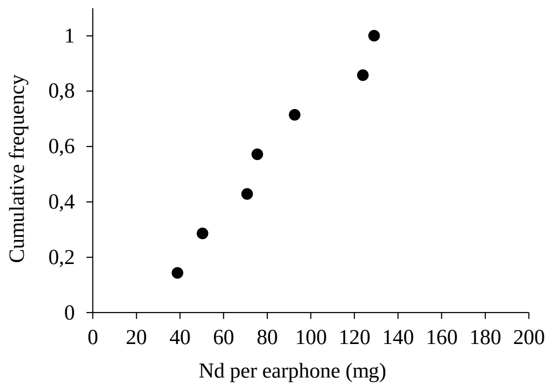

Figure 8 – Cumulative frequency distribution of Nd content of earphones (data from Westphal and Kuchta, 2013)

As mentioned above, in a classical probabilistic approach, one option could be to fit a uniform probability distribution to these data (Figure 9).

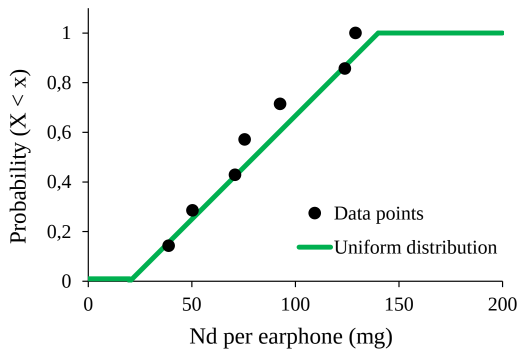

Figure 9 – Fit of the Nd data with a uniform distribution (data from Westphal and Kuchta, 2013)

But these data are scarce and therefore selecting just one distribution introduces a bias in the analysis. Figure 10 shows an alternative, where the data are "bounded" by the family of probability distributions of Figure 6.

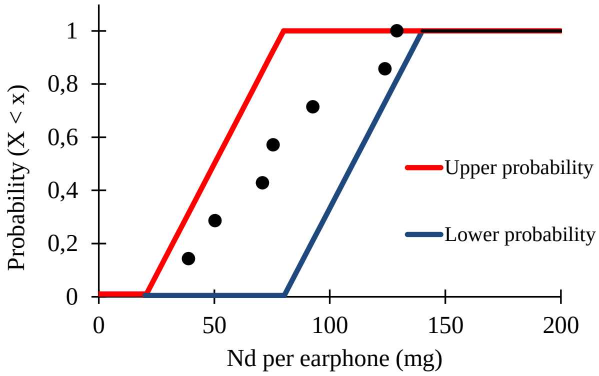

Figure 10 – Upper and lower bounds of probability for the proposal: "Nd-content in earphones is lower than a certain value x"

As will be shown in Section II, such information can be "propagated" in forward modelling (e.g., in life-cycle analysis; Clavreul et al., 2013) or used for data reconciliation in material flow analysis (MFA; Dubois et al., 2014).

Another set of tools, which refers to the general framework of imprecise probabilities of Walley (1991), is probability-boxes (Beer et al., 2013, Ferson et al., 2003). For the data shown in Table 1, one could define the upper and lower bounds of the family of probability distributions based on parametric distributions. Figure 11 shows such a probability box where the upper and lower limits are defined by two triangular distributions (with resp. support [5, 120 mg]; mode = 50 mg and support [10, 180 mg]; mode = 80 mg). The limits could also have been defined using, e.g., normal distributions. However, the latter present the disadvantage of being defined over ± ∞, which can be inconvenient regarding outlier values.

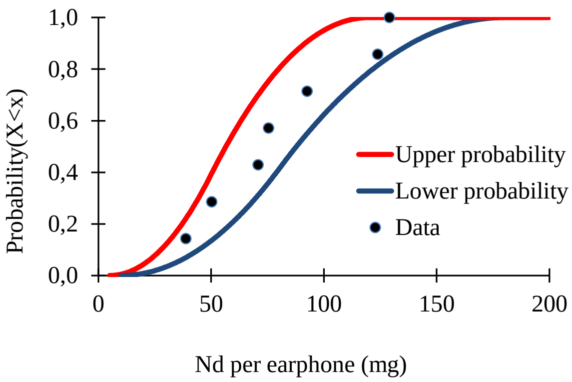

Figure 11 – Probability box where the upper and lower bounds are defined by triangular probability distributions

I.3 "Rich" information regarding Nd concentrations in earphones

Figure 12 illustrates the (hypothetical) situation where there are many measurements of Nd concentrations in earphones. In this case, a single probability distribution can be fitted (using e.g. least-squares) to the data (in this case a normal distribution with average 90 mg and standard deviation 20 mg). The upper and lower probability bounds of Figure 10 are indicated for comparison.

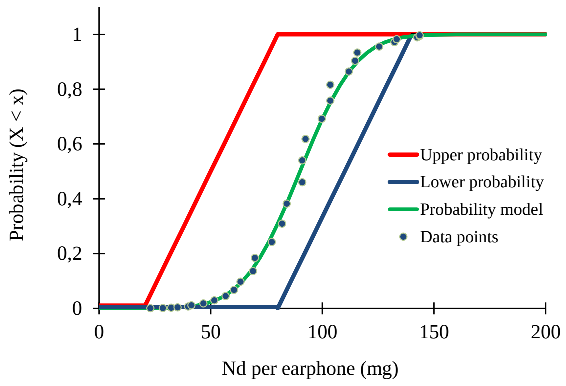

Figure 12 – Least-squares best fit of a normal probability distribution (in green) to a cumulative relative frequency diagram of Nd concentrations in earphone data (black dots; hypothetical data)

II Uncertainty propagation

As most readers will be familiar with Monte Carlo uncertainty propagation, the following illustration will refer largely to the latter. We consider a "Model", noted "M", which is a function of a certain number of parameters, some of which may be constants (K_1, K_2, …, K_m), while others are represented by single (due to "rich information" availability) cumulative probability distributions (P_1, P_2, …, P_n). We are interested in the uncertainty (variability) of $M(K_1, K_2, \ldots K_m, P_1, P_2, \ldots P_n)$. The Monte Carlo method consists in:

(i) Generating "n" random numbers a_1, a_2, … a_n

(ii) Selecting the values p_1, p_2, …, p_n of, resp., P_1, P_2, … P_n, such that: $P_1(X<p_1=a_1)$, $P_2(X<p_2=a_2)$, etc.

(iii) Calculating $M(K_1, K_2, \ldots K_m, p_1, p_2, \ldots p_n)$

(iv) Going back to step (i) and iterating "i" times

Number "i" is selected sufficiently large so as to yield a "smooth" cumulative distribution for $M(K_1, K_2, \ldots K_m, P_1, P_2, \ldots P_n)$. This distribution is obtained by ranking all calculated values in increasing order, assigning a relative frequency of 1/i to each value and then adding up the frequencies.

In the case where certain parameters are based on incomplete/imprecise data, while others can be represented by single probability distributions, methods that allow "joint" propagation of stochastic and epistemic uncertainties are required. One such method is the "independent random set" (IRS) method of Baudrit et al. (2006). We consider a "Model" noted "M", which is a function of a number of parameters, some of which are considered constants (K_1, K_2, …, K_m) while others are represented by single cumulative probability distributions (P_1, P_2, …, P_n) and yet others by possibility distributions (F_1, F_2, …F_q). The IRS methods consists in:

(i) Generating n random numbers a_1, a_2, …, a_n and q random numbers b_1, b_2, …, b_q
(ii) Selecting the values p_1, p_2, …, p_n of, resp., P_1, P_2, … P_n, such that: $P_1(X<p_1=a_1)$, $P_2(X<p_2=a_2)$, etc.
(iii) Selecting the intervals f_1, f_2, …, f_q, of F_1, F_2, …, F_q, at resp. possibility levels b_1, b_2, …, b_q (i.e., alpha-cuts; Dubois & Prade, 1988)
(iv) Calculating the min and max values of $M(K_1, K_2, …K_m, p_1, p_2, …p_n, f_1, f_2, …, f_q)$
(v) Going back to step (i) and iterating "i" times.

The upper and lower bounds of the family of probability distributions $P(X<x)$ are obtained by ranking all min values in increasing order, all max values in increasing order, and assigning a relative frequency of 1/i to each min-max pair and then adding up the frequencies.

These methods are illustrated graphically in Figure 13, where they are applied for estimating the risk of leakage from leach pads used for the extraction of gold in the mining industry. In leach pads, ore is heaped onto barrier materials (clays, geomembranes, …) and leached with aggressive (e.g. acid) solutions to mobilize the economic metal (in this case gold). Leakage through the barrier system (estimated here in liters per hectare per day) represents a risk for groundwater resources. In the calculation result in Figure 13, the Monte Carlo curve fits precisely between the upper and lower bounds obtained from the IRS calculation, because the same shapes were assumed for the possibility distributions as for the probability density functions in the Monte Carlo calculation. Note that in belief theory (Shafer, 1976), the upper probability bound is termed "Plausibility", while the lower probability bound is termed "Belief".

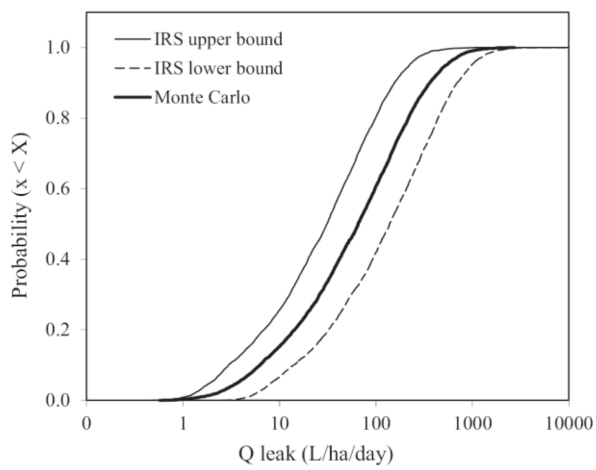

Figure 13 – Example of comparison between IRS and Monte Carlo calculations (Guyonnet & Touze-Foltz, 2011)

In order to obtain a single indicator of the probability of lying below a certain threshold, Dubois & Guyonnet (2011) propose to compute a "confidence index" as a weighted average of the upper and lower bounds of Figure 13. The choice of the weight is subjective and is selected to reflect the decision-maker's level of "aversion to risk". However, this subjectivity is introduced at the final decision-making stage, which is more easily justified than to arbitrarily select single probability distributions at the modeling stage, despite incomplete information. This confidence index is illustrated in Figure 14, where a weight of 1/3 is assigned to the "optimistic" distribution (here the upper bound) while a weight of 2/3 is assigned to the "pessimistic" distribution (the lower bound). It would seem "natural" to assign more weight to the "pessimistic" distribution, but without completely ignoring the "optimistic" one.

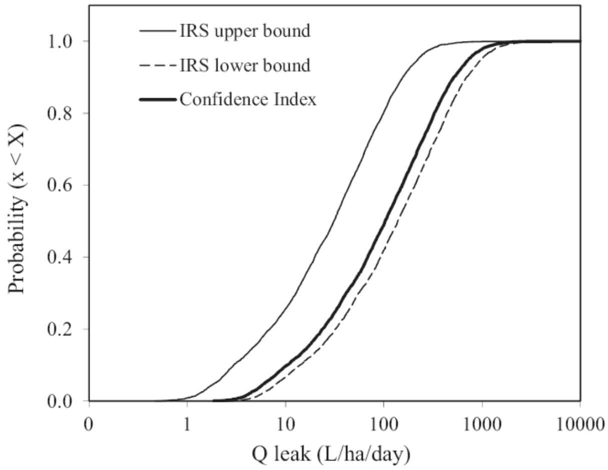

Figure 14 – Illustration of the confidence index for a weight of 1/3 (Guyonnet & Touze-Foltz, 2011)

Contexts of use, application fields	-> contexts (e.g., environmental, economic, social assessment) -> which types of stakeholder questions are concerned? -> link to published studies that implement the method

Uncertainty representation and propagation can be applied to:

- Measurement errors
- Risk analysis
- Material flow analysis (MFA – see the factSheet)
- Life cycle assessment (LCA – see the factsheet)
- Cost-benefit analysis
- Etc.

The context and applicability is very transverse, as variability and imprecision/incompleteness affect nearly all areas of mineral intelligence capacity.

Input parameters	-> which parameters are needed to run the method

Input parameters depend on the selected method of uncertainty representation. For a purely stochastic approach to uncertainties, moments describing the probability distributiuons are required (e.g., averages and standard deviations in the case of Gaussian probability distributions). In the case of possibility distributions, based on expert knowledge or scarce information, the supports and cores of the distributions are required.

Type(s) of related input data or knowledge needed and their possible source(s)	-> which types of data are needed to run the method, from which sources could they come… -> could be qualitative data or quantitative data, and also tacit knowledge, hybrid, etc.

As above. Selected method should be tailored to the type of information at hand (expert judgment, scarce data, rich data, etc.).

Model used (if any, geological mathematical, heuristic…)	-> e.g., geological model for mapping -> e.g., mathematical model such as mass balancing, matrix inversion, can be stepwise such as agent -based models, dynamic including time or quasidynamic specifying time series… -> can also be a scenario

The approach is particularly well suited to models that are in a “closed form”; i.e. a mathematical equation. Numerical models may also be amenable but may be costly in terms of calculation times.

System and/or parameters considered	-> **the system can be described by its boundaries**. These can refer to a geographic location, like a country, or a city, the time period involved, products, materials, processes etc. involved, like flows and stocks of copper, or the cradle-to-grave chain of a cell phone, or the car fleet, or the construction sector, or the whole economy… -> **parameters** could possibly refer to geographic co-ordinates, scale, commodities considered, genesis of ore deposits and others…

The approach is generic and therefore suited to a wide variety of systems.

Time / Space / Resolution /Accuracy / Plausibility…	-> to which spatio-temporal domain it applies, with which resolution and/or accuracy (e.g., near future, EU 28, 1 year, country/regional/local level…) -> for foresight methods can also be plausibility, legitimacy and credibility…

Case by case basis.

Indicators / Outputs / Units	-> this refers to what the method is actually meant for. Units are an important part but that is most of the time not sufficient to express the meaning. For example, **the indicators used in LCA express the cradle-to-grave environmental impacts of a product or service**. This can be expressed in kg CO_2-equivalent. But also in €. Or in millipoints. Or in m^2year land use. -> for foresight methods the outputs are products or processes

Case by case basis.

Treatment of uncertainty, verification, validation	-> evaluation of the uncertainty related to this method, how it can be calculated/estimated

This is the object of the FactSheet.

References cited	

Baudrit, C., Dubois, D., Guyonnet, D. (2006). Joint propagation and exploitation of probabilistic and possibilistic information in risk assessment models. IEEE Transactions on Fuzzy Systems, vol.14, No.5, pp.593-608. **doi 10.1109/TFUZZ.2006.876720**.

Beer, M., Ferson, S., Kreinovich, V. (2013). Imprecise probabilities in engineering analyses. Mechanical Systems and Signal Processing, 37, 4–29. **doi.org/10.1016/j.ymssp.2013.01.024**

Clavreul, J., Guyonnet, D., Tonini, D., Christensen, T. H. (2013). Stochastic and epistemic uncertainty propagation in LCA. International Journal of Life Cycle Assessment, 18, 1393-1403. **http://dx.doi.org/10.1016/j.wasman.2012.07.008**

Dubois, D., and Prade, H. (1988). Possibility theory. New York Plenum Press, 263 pp. **ISBN 10: 0306425203**

Dubois, D., Guyonnet, D. (2011). Risk-informed decision-making in the presence of epistemic uncertainty. International Journal of General Systems, 40(2), 145-167.
doi: 10.1080/03081079.2010.506179

Ferson, S., Kreinovich, V., Ginzburg,L., Myers, D.S., Sentz, K. (2003). Constructing probability boxes and Dempster–Shafer structures. Tech. Rep. SAND2002-4015, Sandia National Laboratories, Albuquerque, NM.

Ferson, S. (1996). What Monte Carlo methods cannot do. Human and Environmental Risk Assessment, 2, 990-1007.

Ferson, S. and Ginzburg, L. (1996). Different methods are needed to propagate ignorance and variability. Reliability Engineering and System Safety, 54, 133-144.
https://doi.org/10.1016/S0951-8320(96)00071-3

Guyonnet, D., Touze-Foltz, N. (2014). Assessing the risk of leakage from heap leach pads used in mining operations. The Extractive Industries and Society, 1(1), 96-103.

Jaynes, E. T. (1988). The Relation of Bayesian and Maximum Entropy Methods. In: Maximum-Entropy and Bayesian Methods in Science and Engineering, Vol. 1, Kluwer Academic Publishers, p. 25-29.
ISBN 90-277-2793-7

Shafer, G. (1976). A mathematical theory of evidence. Princeton University Press.
ISBN: 9780691100425

van Beers, D., R. van Berkel, and T. E. Graedel. (2005). The Application of Material Flow Analysis for the Evaluation of the Recovery Potential of Secondary Metals in Australia. In: 4th Australian LCA Conference. Sydney, Australia.

Walley, P. (1991). Statistical reasoning with imprecise probabilities. Chapman & Hall, London,1991.
ISBN-10: 0412286602

Westphal, L., Kuchta, K. (2013). Permanent magnets from small waste electrical and electronic equipment. In: Proceedings Sardinia 2013, Fourteenth International Waste Management and Landfill Symposium. Cagliari, Italy: CISA Publisher.

Zadeh, L. (1978). Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets and Systems, 1, 3-28.
doi.org/10.1016/S0165-0114(99)80004-9

Related methods	-> List of comparable methods, their particularities… -> link to one or several other existing fact sheet(s)

Related methods are:

- Sensitivity analysis
- Scenario analysis
- Geostatistics
- Probability boxes
- 2D Monte Carlo
- Bayesian methods
- …

Related factSheets are (among others):

- FactSheet: '**Life Cycle Assessment**'
- FactSheet: '**Using dynamic MFA or system dynamic modelling**'
- FactSheet: '**Cost Benefit Analysis**'
- FactSheet: '**Input Output Analysis**'

Some examples of operational tools (CAUTION, this list is not exhaustive)	-> e.g., software… Only give a listing and a reference (publication, website/page…) -> **should be provided only if ALL main actors are properly cited**

Examples of tools available either free or commercially:

- Commercially available: Purely stochastic risk analysis: @RISK; http://www.palisade.com/risk/

- Public domain: joint stochastic/epistemic propagation: HYRISK; https://cran.r-project.org/web/packages/HYRISK/index.html

Key relevant contacts	-> list of relevant **types** of organisations that could provide further expertise and help with the methods described above.

Purely stochastic uncertainty analysis is routinely performed by consulting companies. Joint propagation methods are currently developed in research institutes. A reference institute for such methods is: **https://www.irit.fr**

Glossary of acronyms /abbreviations used	-> Definition