METALS ENVIRONMENTAL RISK ASSESSMENT GUIDANCE

FACT SHEET

MERAG 07

UNCERTAINTY ANALYSIS

January 2007
Check you have the most recent fact sheet by visiting www.metalsriskassessment.org

defra
Department for Environment Food and Rural Affairs

euras

EuroMetaux

ICMM
International Council on Mining & Metals

CONTENTS

The content of the MERAG fact sheets reflect the experiences and recent progress made with environmental risk assessment methods, concepts and methodologies used in Chemicals Management programs and Environmental Quality Standard setting (soil, water, sediments…) for metals. Since science keeps evolving these fact sheets will be update on a regular basis to take into account new developments. To be sure you have the most recent fact sheet on the current subject check our website: www.metalsriskassessment.org.

1. INTRODUCTION

Any risk assessment always carries uncertainty with it. The current risk assessment schemes are largely deterministic and uncertainty issues are most often accounted for by means of worst-case assumptions and assessment factors which are implicitly embedded in calculation schemes and rules. However, combining risk and uncertainty into a single measure makes the risk assessment not transparent to a risk manager or decision-maker as to how conservative or realistic the risk outcomes actually are. There is, therefore, a need for a proper and transparent treatment of sources of variability and uncertainty during the risk assessment process. One way to classify uncertainty is to differentiate it into uncertainty due to variability and uncertainty due to limited knowledge (van Asselt, 2000; Walker et al., 2003). The terminology used in the risk assessment jargon to describe these uncertainties is variability and uncertainty.

Data and models used in environmental exposure, ecological effects and risk assessment are also characterized by variability and uncertainty and this has been recognized by both in both the academic literature (e.g. Burmaster, 1997; Campbell *et al.*, 2000; Cullen & Frey, 1999; Warren-Hicks & Moore, 1995; Hart, 2001; Posthuma et al., 2002; Vose, 1996) and in a regulatory context (e.g. Jager *et al.*, 1997; 1998; 2000; 2001; EPA, 1997; 1999; 2001; ECOFRAM, 1999). The National Academy of Sciences (NRC, 1983) has recommended that the distinction between variability and uncertainty should be maintained rigorously at the level of individual components of a risk assessment (e.g. emissions characterization, exposure assessment) as well as at the level of an integrated risk assessment. A workshop sponsored by the US Environmental Protection Agency provided recommendations regarding the use of two-dimensional simulations, which were incorporated into a 1997 agency policy document (EPA, 1997).

Variability represents inherent heterogeneity or diversity in a well-characterized population. Fundamentally a property of nature, variability is not reducible through further measurement or study. The most well known sources of variability in environmental risk assessment are the temporal and spatial variations of the environmental concentration (captured in the Exposure Concentration Distribution (ECD)) and the inter- and intra-species sensitivity of a toxicant (captured in respectively Species Sensitivity Distribution (SSD) and a dose-response relationship). Uncertainty represents ignorance and lack of complete information about poorly characterised phenomena or models (e.g. sampling or measurement error), and can partly be reduced through further research (Cullen & Frey, 1999). A well known source of uncertainty is the sampling uncertainty. This is uncertainty due to limits on sample size which is evitable given that one needs an infinite number of samples to obtain a correct estimate of a parameter. Sampling uncertainty is sometimes calculated as a confidence interval for example on the HC_5 estimate of a SSD.

Box 1: Variability versus uncertainty in risk assessment

For uncertainty analysis, it is helpful to verify whether the parameter under consideration is characterized by variability, uncertainty or both. This verification is based on the definition of variability and uncertainty and the scope of the assessment. An overview of sources of uncertainty and variability in risk assessments of metals is given in the Table below.

Table 1: Overview sources of uncertainty and variability

	Variability: - Not reducible through further research - Probability distribution represents real variations actually occurring	**Uncertainty**: - Reducible through further research - True value is somewhere in the probability distribution
Exposure	• Temporal variability of emission, effluent concentration, flow… • Spatial variability of soil/sediment/aquatic characteristics, background concentration,… in regional assessment	• E-USES or BLM model structure uncertainty • Uncertainty of physico-chemical properties, partition coefficients, removal rates,… • Spatial variability of soil/sediment/aquatic characteristics, background concentration,… in local assessment
Effects	• Intra-species sensitivity • Inter- and intra-laboratory variability • Endpoint differences • Spatial (and temporal) variability of physico-chemical characteristics determining bio-availability • Inter-species sensitivity	• Probability distribution uncertainty (e.g. threshold versus non-threshold distribution) • Diversity & representativeness • Sampling uncertainty • Lab to field extrapolation

A metal effluent concentration of a local site is, for example, characterized by temporal variability and by sampling uncertainty (i.e. the fact that instead of an infinite number of effluent concentration measurements, only monthly measurements are available). The aquatic metal concentration in Wales, as another example, is mainly be characterized by spatial variability. The sampling uncertainty on the regional aquatic metal measurements can be considered negligible if the sample size is very large. The measurement error or uncertainty can also be considered as negligible compared to the spatial variability. Sometimes, a parameter can be considered as variable or uncertain depending on the scope/scale of the assessment. The concentration of suspended solids, for example, is on a regional level characterized by spatial variability. The same parameter cannot be characterized by spatial variability on a local site. However, if the concentration of suspended solids is not known at the local site, the spatial variability probability distribution on a regional level can serve as a surrogate uncertainty probability distribution on a local level.

If a parameter is found to be variable and uncertain, several methods can be used to estimate both uncertainty and variability at the same time in a two-dimensional analysis: bootstrapping, Bayesian analysis, classical methods,… More information on these methods can be found in literature (Cullen & Frey, 1999), (Davison & Hinkley, 1997), (Efron & Tibshirani, 1993).

Uncertainty and risk should ideally be assessed separately. Figure 1 illustrates that risk and uncertainty are two independent concepts. A risk assessment may, for example, result in a large estimated risk but this estimated risk can either be very uncertain or the estimated risk may only be characterized by limited uncertainty. The same holds true for a small estimated risk.

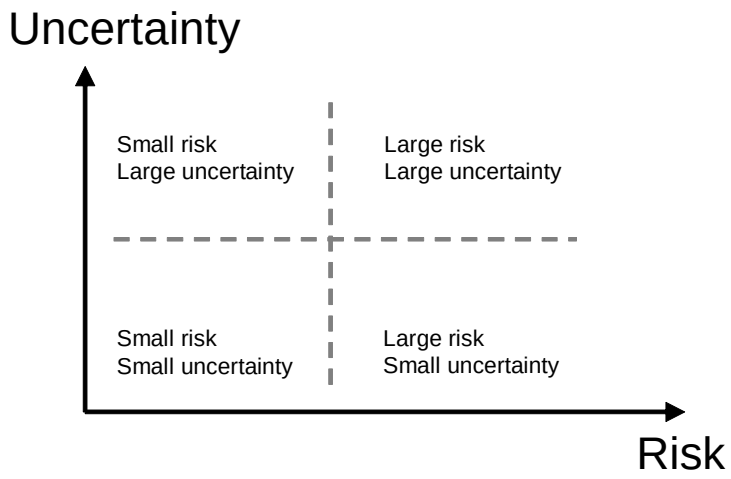

Figure 1: Illustration showing that risk and uncertainty are two independent concepts

Methods for characterizing risk and its uncertainty should not be conditioned by data availability. Probabilistic methods are often introduced as a way to better quantify uncertainty and variability in risk assessment. However, uncertainty can also be estimated in data poor conditions[1] – as is often the case in current deterministic assessments. This requires some more pragmatic tools.

In case of data poor metals and lack of prior knowledge on distribution shape of variability and uncertainty, the deterministic approach may be the only valid way forward. Since, in a conventional deterministic risk assessment, the different layers of conservatisms used in the risk assessment are eventually “hidden” in the risk quotient estimates, it is recommended to transparently report the different sources of uncertainty and where possible and useful to provide a quantitative estimate of the uncertainty (e.g. through a scenario analysis, see further).

In case of deterministic risk (i.e. when risk characterization ratio becomes larger than one) or for data rich metals, probabilistic techniques could be used. Probabilistic techniques may require the collection of additional data (e.g. more than 10 toxicity tests) but usually result in more realistic and differentiated risk statements (including both quantitative and qualitative uncertainty estimates). Again, both the uncertainty and variability are not “hidden” but are explicitly communicated in the risk characterization to the risk manager. Probabilistic risk assessment can therefore be considered as a precautionary-driven approach because in this way, issues of (pre)caution due to uncertainty are explicitly transferred to the risk management phase (Verdonck et al., 2005). To identify the most critical parameters driving the risk, it is recommended to perform a sensitivity analysis as a first step in the probabilistic risk framework.

[1] Data poor conditions are considered to contain enough data to calculate some exposure and effects point estimate. If no data on exposure or effects are available, no uncertainty analysis and risk characterization can be carried out.

2. DIFFERENTIATING UNCERTAINTY FOR DATA-POOR AND DATA-RICH METALS

A first prerequisite is to improve the transparency of the uncertainty. For this, uncertainty can be classified into several classes. Several guidances have been developed in response to the notion that in the daily practice of science for policy, there is a pressing need for guidance in assessing and communicating uncertainties (van der Sluijs et al., 2003; Krayer von Kraus, 2005b). The RIVM guidance recognizes that this need extends beyond the quantitative assessment of uncertainties in model results per se, and focuses therefore on the entire process of environmental assessment, running from problem framing towards reporting the results of the study (van der Sluijs et al., 2003). Arguably, with the development and implementation of the guidance, RIVM sets a best practice standard in environmental management.

The uncertainty matrix is in this regard an useful aid in making an inventory of where ('location') the most (policy) relevant uncertainties are expected, and how they can be characterized in terms of a number of uncertainty features (Walker et al., 2003; van der Sluijs et al., 2003; 2004; Krayer von Kraus, 2005b). In the draft TGD for REACH (RIP, 2005), a check list is proposed to systematically check and list the different sources of uncertainty and variability for each step in the risk assessment procedure. Uncertainty matrices and check lists allow the risk assessor and manager to have a broader perspective on uncertainty instead of only focusing on the known or quantifiable sources. Recognized but unquantifiable uncertainties are in this way made more transparent.

As exposure concentrations (PEC) and the effects thresholds (PNEC) are characterized by both variability and uncertainty such variables are called second-order variables and represented by probability distributions in two dimensions. For each percentile of the variability distribution, an uncertainty or confidence interval can be estimated (i.e. the uncertainty distribution). In **Figure 2**, a variability distribution is represented as a cumulative distribution function. The uncertainty distribution can also be represented by a cumulative or density distribution function; however, for communication purposes, it is often represented by a 90% uncertainty or confidence interval or band (see grey bands in Figure 2). The worst-case exposure estimate (i.e. the PEC) used in actual risk assessments can be seen as an upper percentile of both its temporal and/or spatial variability and its uncertainty. The actual worst-case effects or toxicity estimate (i.e. the PNEC) can be seen as a lower percentile of both its inter-species, and other forms of variability, and its uncertainty.

When insufficient data and/or knowledge do not allow estimating these distributions, assessment Factors (AF) and worst-case assumptions are used to estimate the uncertainty and variability (as illustrated by the arrows at the bottom of Figure 2). The actual worst-case exposure estimate (i.e. the PEC) can then be seen as a mean exposure estimate multiplied with extrapolation factors due to worst-case assumptions both from variability and uncertainty. The actual worst-case effects or toxicity estimate (i.e. the PNEC) can be seen as a mean effects or toxicity estimate divided by assessment factors and worst-case factors both from variability and uncertainty. This can also mathematically be formulated as:

$$WorstCaseExposure = MeanExposure \cdot FactorExp_{worstcase,\mathrm{var}} \cdot FactorExp_{worstcase,unc}$$

$$WorstCaseToxicity = \frac{MeanToxicity}{\left(AF_{\mathrm{var}} \cdot FactorTox_{worstcase,\mathrm{var}}\right) \cdot \left(AF_{unc} \cdot FactorTox_{worstcase,unc}\right)}$$

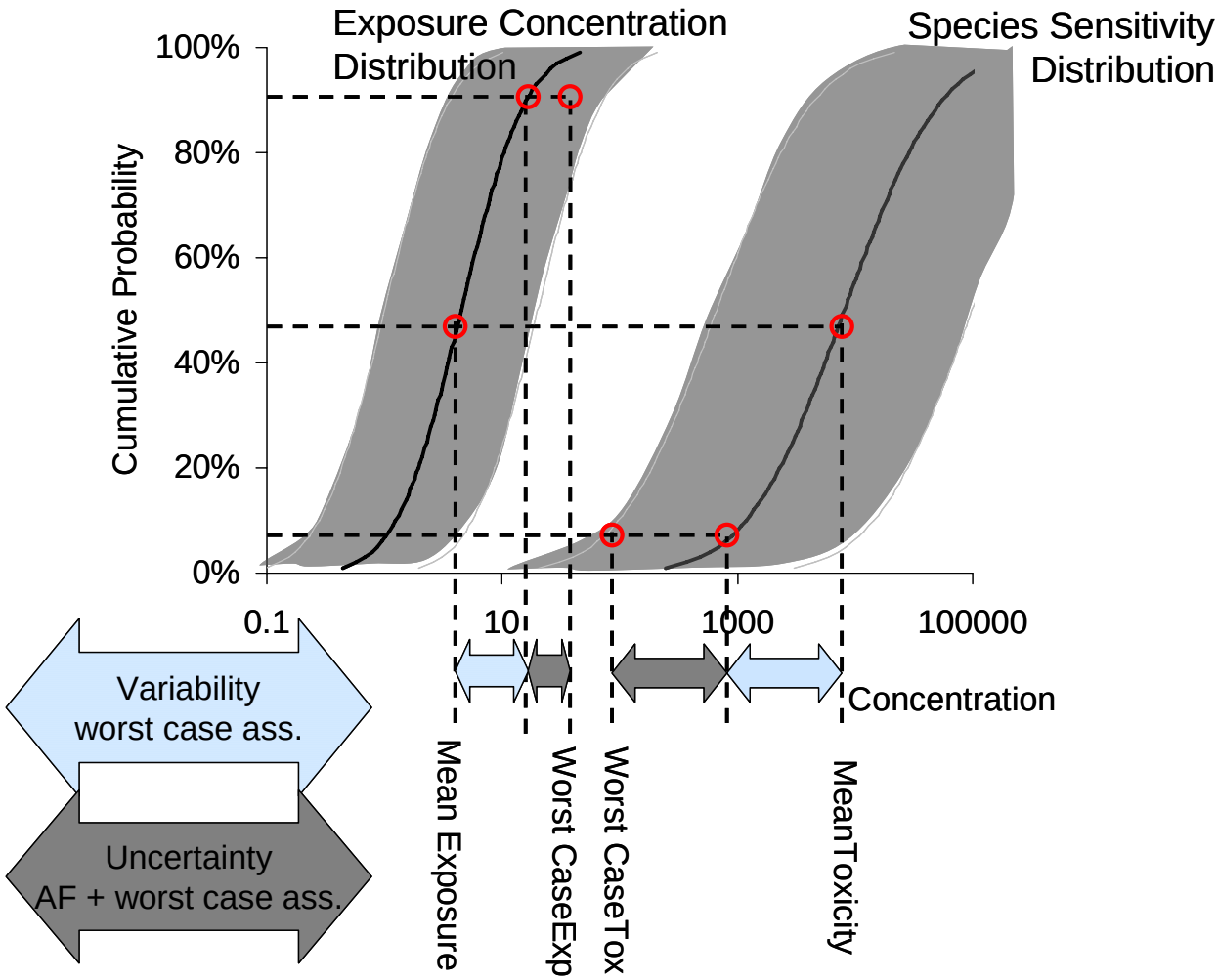

Figure 2: Top: Environmental cumulative probability distributions describing variability in exposure (Exposure Concentration Distribution) and effects/toxicity (Species Sensitivity Distribution or SSD) with grey bands describing 90% uncertainty. Bottom: assessment factors/worst case assumptions describing variability or uncertainty in environmental risk assessment (AF: Assessment Factors)

3. UNCERTAINTY ANALYSIS FOR DATA-POOR METALS

The reflex on uncertainty to strive for certainty has led to the build-up of conservative measures (worst case assumptions and assessment factors) in a risk assessment context. However, application of worst case assumptions and assessment factors is often not transparent or scientifically appropriate and may lead for metals to unrealistic risk outcomes (e.g. PNECs below the background).

An uncertainty matrix or a check list allows already the risk assessor and manager to locate the most important sources of uncertainties and how they can be characterized. A crude scenario-analysis can be subsequently conducted to estimate and gain insight in realistic order of magnitudes of uncertainty. For this, a distinction needs to be made between measures taken in the risk assessment process due to variability (inherent, irreducible environmental variations), such as taking the 5^{th} percentile of SSD (due to inter-species sensitivity) or taking the 90^{th} percentile of a local effluent concentration (due to temporal variability), and measures due to uncertainty such as applying assessment factor (due to lab-to-field or model uncertainty). It is recommended to quantify the degree of uncertainty introduced at each level of the risk assessment process. This can be done by conducting the risk assessment for the following two or three scenarios:

1. The **reasonable worst-case scenario** accounts for all (realistic) worst-case assumptions and assessment factors caused by sources of both variability and uncertainty. This scenario mostly considers parameters and assumptions towards the lower end conditions (on effect side) or higher end conditions (on exposure side).
2. The **typical scenario** accounts for the worst-case assumptions and assessment factors caused by sources of variability only. This scenario considers worst-case conditions only for parameters characterized by variability. Averages or medians are mostly taken for parameters characterized by uncertainty.
3. The **average scenario** does not account for sources of variability and uncertainty. This scenario is not necessarily sufficiently protective for the environment and should therefore not always be considered.

In Table 1 the mathematical expression of the calculation of the risk characterization ratio (RCR) is given for the different scenarios.

Table 1: Risk Characterization Ratio (RCR) formula for several scenarios in data-poor conditions

Scenario	Formula Risk Characterization Ratio (RCR)
Reasonable worst-case scenario (accounts for both variability and uncertainty)	$RCR = \frac{MeanExposure \cdot FactorExp_{worstcase,\mathrm{var}} \cdot FactorExp_{worstcase,unc}}{\frac{MeanToxicity}{AF_{\mathrm{var}} \cdot AF_{unc} \cdot FactorTox_{worstcase,\mathrm{var}} \cdot FactorTox_{worstcase,unc}}}$
Typical scenario (accounts for only variability)	$RCR = \frac{MeanExposure \cdot FactorExp_{worstcase,\mathrm{var}}}{\frac{MeanToxicity}{AF_{\mathrm{var}} \cdot FactorTox_{worstcase,\mathrm{var}}}}$
Average scenario	$RCR = \frac{MeanExposure}{MeanToxicity}$

The difference in Risk Characterization Ratios (RCRs) between the reasonable worst-case and typical scenario can thus be considered as a measure for uncertainty. The difference in RCRs between the typical and average scenario can be considered as a measure for variability. The average scenario is not necessarily sufficiently protective for the environment but can be useful to obtain a quantitative estimate of the variability.

Figure 3 provides a visual representation of the level of uncertainty introduced when comparing the outcome of the exposure and effect assessment for both scenarios as mentioned above. The build-up of uncertainty can be quantified by calculating the overall uncertainty factor (ratio $PEC_{scenario\ 1}/PEC_{scenario\ 2}$; ratio $PNEC_{scenario\ 2}/PNEC_{scenario\ 1}$) after the last step in the PEC and PNEC calculation.

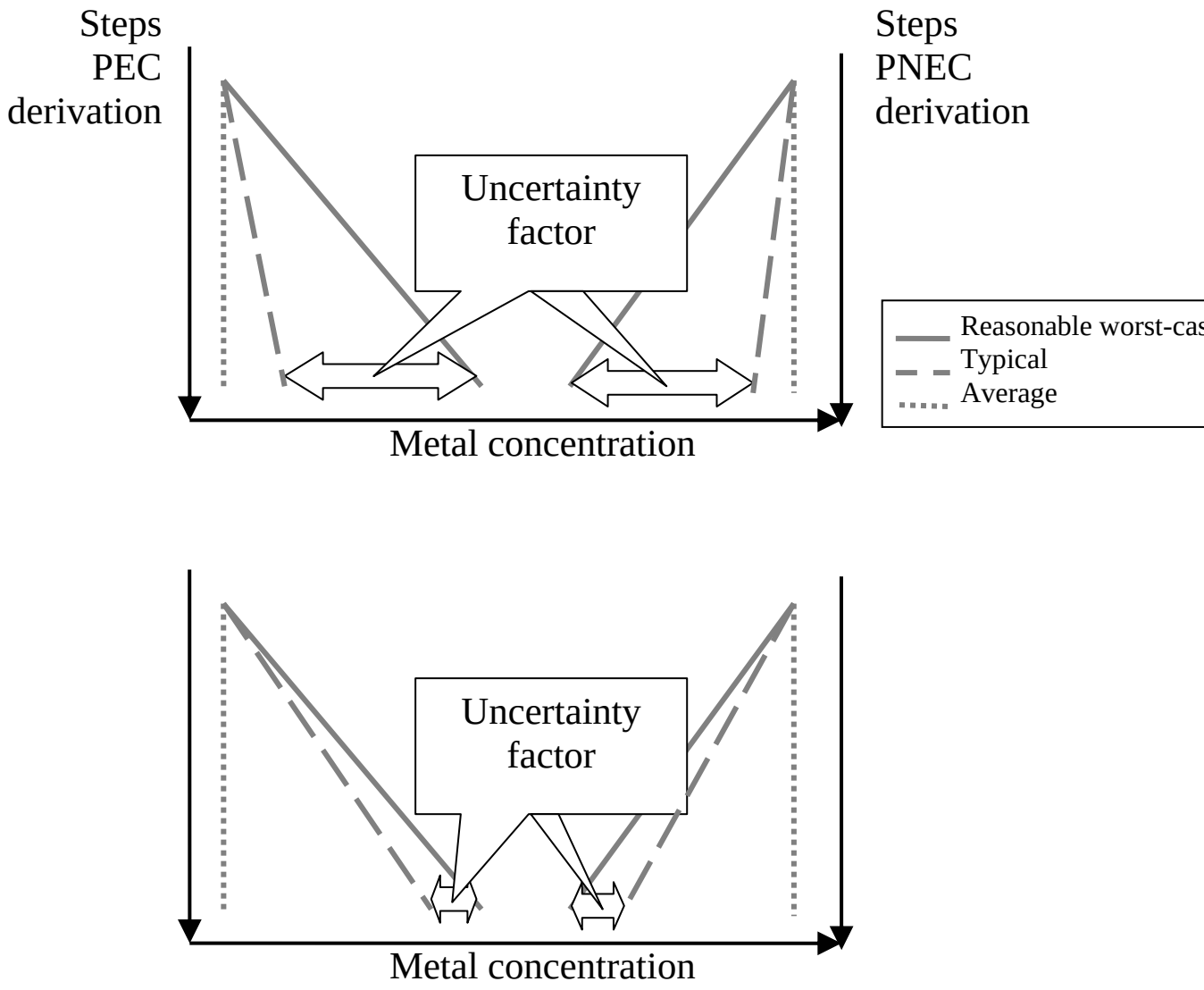

Figure 3: Conceptual scheme of increasing level of uncertainty when going through all calculation steps in the exposure (left) and effects (right) assessment. The uncertainty factor can be interpreted as the MOS (Margin of Safety) in human health risk assessments. The top graph illustrates an assessment with more uncertainty compared to the lower graph illustrating an assessment with less uncertainty.

An indicative overview of the assumptions to be considered in the different scenarios for the exposure, effects assessment and risk characterization is given in Table 2. In the reasonable worst-case scenario, minima or low end values (e.g. 10th percentiles are typically taken in the effects assessment and high end values (e.g. 90th percentiles) or maxima are taken in the exposure assessment. In addition, assessment factors are often implemented. Contrary to the typical scenario, averages or medians of uncertainty variations are taken in both effect/exposure assessment and no additional assessment factors are used in the reasonable scenario. For some parameters as the Kd-value, the minimum or 10th percentile is a worst case estimate for one compartment but is, at the same time, a best case scenario for another compartment. For these parameters, both a minimum or 10th percentile and a maximum or 90th percentile could be used in the assessment.

The uncertainty analysis methodology can be conducted for both a local and regional risk assessment. Obviously, the steps given in Table 2 will differ for a local and regional assessment but the principles of calculating the level of uncertainty remain the same. Spatial variability is not relevant in a local assessment as site-specific conditions apply.

Table 2: Indicative overview of uncertainty analysis in regional/local deterministic risk assessment

Step		**Source of uncertainty or variability**	**Uncertainty analysis in data-poor metal risk assessment**		
			uncertainty ⟷	**variability** ⟷	
			Reasonable worst-case scenario	Typical scenario	Average scenario
Effects assessment	1. Data selection	Intra-species sensitivity	NOEC	EC_x	EC_x
	2. Data aggregation	Inter- and intra-laboratory variability	Geometric mean*	Geometric mean	Geometric mean
		Endpoint differences	Lowest is taken	Lowest is taken	Median is taken
	3. Bio-availability normalization	Spatial (and temporal) variability of physico-chemical characteristics	to low end value (e.g. 10P) conditions	to average conditions and lowest BLM model	to average conditions and average BLM model
	4. Selection distribution**	Model/ distribution uncertainty	Lognormal/loglogistic is standard***	Take best fitting distribution	Take best fitting distribution
	5. Derivation protection level	Inter-species sensitivity	HC_5 ** Assessment factor between 10-1,000	HC_5 ** Assessment factor between 10-1,000	HC_5 ** Assessment factor between 10-1,000
	6. Derivation PNEC	Overall quality data	Assessment factor between 1-5	No assessment factor	No assessment factor
		Diversity & representativeness	Assessment factor between 1-5		
		Sampling uncertainty	Assessment factor between 1-5		

Step		Source of uncertainty or variability	Uncertainty analysis in data-poor metal risk assessment		
			uncertainty ↔ variability		
			Reasonable worst-case scenario	Typical scenario	Average scenario
		Lab to field extrapolation	Assessment factor between 1-5		
Exposure modelling	Parameter collection	Temporal variability of emission (both local and regional level)	Average is taken*	Average is taken	Average is taken
		Temporal variability of effluent concentration (typically for local assessment)	High end value (e.g. 90P) is taken	High end value (e.g. 90 P) after elimination of outliers	Average is taken
		… physico-chemical properties, partition coefficients, removal rates	Averages or percentiles are taken	Average/ median is taken	Average/ median is taken
		Other parameters	Reasonable worst case	Average is taken	Average is taken
Exposure monitoring	1. Data aggregation	Temporal variability	High end value (e.g. 90th P) is taken for each site	High end value (e.g. 90th P) is taken for each site	Median is taken for each site
	2. Derivation PEC	Spatial variability	Mean is taken*	Mean is taken	Mean is taken
	3. Derivation PECadd if needed	Sampling uncertainty (usually small for large data sets)	Not considered	Not considered	Not considered
		Spatial variability of background concentration	Subtract minimum	Subtract site specific value or in absence, average	Subtract average
charac terizati			Reasonable worst-case risk quotient	Typical risk quotient	Average risk quotient

Step	Source of uncertainty or variability	Uncertainty analysis in data-poor metal risk assessment		
		uncertainty ⟷		variability ⟷
		Reasonable worst-case scenario	Typical scenario	Average scenario
		Level of uncertainty: difference between reasonable worst-case and typical risk quotient		

* *This is not the most realistic conservative option*

** *In case sufficient data are available (criteria, see TGD)*

*** *This may not always be the most realistic conservative option*

As a reality check, the total amount of uncertainty included, being the difference between the worst case and the typical scenario, could be evaluated/ compared with e.g. the following criteria/aspects. This can be done using a weight of evidence approach to determine the total amount of acceptable uncertainty that could be applied in the final risk characterization phase.

Issues which should be included in this weight of evidence assessment are:

- the overall quality and relevance of the data bases;
- the amount and representativity of the data used;
- the used assessment factors;
- the margin between the toxicity level and the natural background/essentiality levels;

It is recommended that all sources of uncertainty are transparently presented in the risk assessment report.

4. UNCERTAINTY ANALYSIS FOR DATA-RICH METALS

In case a deterministic risk has been observed (i.e. when risk characterization ratio becomes larger than one) or in case the metal of concern is a data-rich metal, probabilistic techniques are recommended as a higher tier. Probabilistic techniques may require the collection of additional data and hence result in additional effort.. The outcome of the risk characterization is often mostly influenced by specific input parameters and assumptions related to both the effects (e.g. how does the selection of a specific SSD, threshold or non-threshold probability distribution, influence the risk characterization?) and exposure assessments (e.g. how does bio-availability affect the risk characterization?). To identify and rank the critical parameters or assumptions that drive the risk characterization and consequently efficiently allocate resources for further data collection, it is recommended to use sensitivity analysis as a first step (see Figure 4). In a next step, the variability and uncertainty of the most sensitive parameters are quantified. Finally, the quantified uncertainty and variability are propagated through the exposure or effects model.

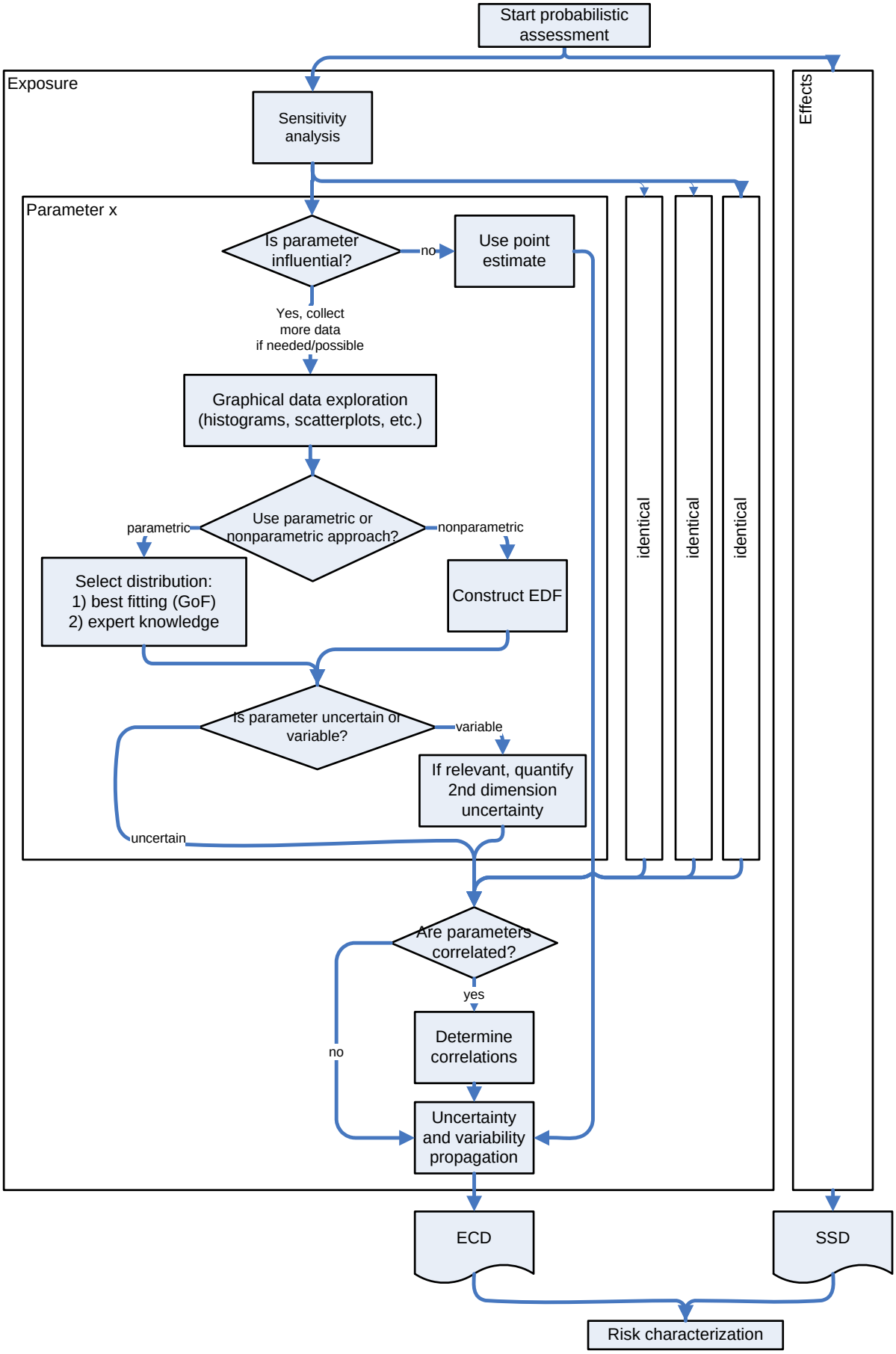

Figure 4: Short overview flowchart uncertainty analysis in probabilistic risk assessment

4.1 Sensitivity analysis

Sensitivity analysis, as it is applied to risk assessment, is an approach to determine which factors in a risk model (specific exposure pathways or making certain assumptions with respect to model parameters) influence risk most strongly. It provides a means of exploring, in a quantitative manner, the effect of a variety of "what-if" scenarios on the risk estimates. The basic approach is to allow for a subset of the input variables to vary within prescribed ranges and to determine how much the model output (usually risk) changes in response to changes in the values for each input variable. Of the several approaches to sensitivity analysis that are available, no single approach will serve as the best analysis for all modelling efforts. The best choice for a particular situation will depend on a number of factors, including the nature and complexity of the model and the resources available. For example, sensitivity ratios can be used where the ratio is equal to the percentage change in output (e.g. risk) divided by the percentage change in input for a specific input variable. Risk estimates are considered most sensitive to input variables that yield the highest ratios. Guidance on how to perform a sensitivity analysis could be consulted in Saltelli et al. (2000) (or in Cullen & Frey (1999)).

Box 2: Local and global sensitivity analysis

Sensitivity ratios can generally be grouped into two categories: local and global. For the local sensitivity analysis, an input variable is changed by a small amount, usually ±5% of the nominal (default) point estimate, and the corresponding change in the model output is observed. For global sensitivity analysis, an input variable is varied across the entire range (plausible minimum and maximum values). If local and global sensitivity results are different, it can be concluded that different exposure variables are dominating risk near the high-end (i.e. extreme tails of the risk distribution) than are dominating risk at the central tendency. This situation is likely to occur when there are nonlinear relationships between an input and output variable.

Sensitivity analysis is beneficial as it helps in deciding whether and where more information is needed to refine the analysis and therefore provide a powerful tool to reduce the uncertainty associated with such assessment. In that respect sensitivity analysis could be used in both point estimates and probabilistic approaches. A sensitivity analysis is particularly useful when applied in a tiered approach in deciding which exposure pathways and assumptions are carried forward from a point estimate risk assessment into a one- or two-dimensional Monte Carlo analysis (definition see further). By identifying the variables that are most important in determining risk, one can also decide whether point estimates, rather than probability distributions, can be used with little consequence to the model output (thereby reducing the level of effort associated with developing probability distributions for all input variables).

4.2 Uncertainty analysis

Once the main uncertainty drivers are identified, the uncertainty of those parameters can be quantified and considered for uncertainty analysis.

4.2.1. Uncertainty characterization of the input parameters

If a parameter is found to be important to the risk outcome, additional data can be collected (if not already available) to characterize its variability and/or uncertainty. A number of steps need to be conducted in order to characterize the uncertainty of an input parameter. A number of data points or expert knowledge is needed to quantify its uncertainty. It can then be checked whether a parametric or nonparametric distribution is more appropriate, whether the parameter of concern is variable, uncertain or both and whether there exist correlations/dependencies between several input parameters. Graphical

plots as histograms and scatterplots can be used to explore the data of the parameter and help in selecting a parametric or nonparametric distribution. A parametric distribution could be selected based on best fitting criteria (such as goodness-of-fit tests) and expert knowledge. The risk assessor may, for example, consider threshold models for essential metals. Or, the risk assessor may consider a truncated distribution (min = 0; max = 14) for characterizing the parameter pH (a parameter that can influence bio-availability). The parametric or nonparametric distribution should then be characterized. Next step is to explore whether the parameter under investigation is considered to be uncertain, variable or both. If a parameter is found to be uncertain and variable at the same time, like for example the effluent concentration of a metal at a local site varies in time but can also be uncertain due to a limited number of samples, its variability and uncertainty could be characterized.

More information on selecting and fitting distributions can be found in US EPA guidance: EPA (2001) and EPA (1999) and other literature: Cullen & Frey (1999), Vose (1996).

Box 3: Parametric or nonparametric distributions

Guidance on the use of parametric or nonparametric approaches can be found in EPA (1999), Cullen and Frey (1999) and Vose (1996).

In case of doubt on selecting a parametric or nonparametric method, both methods are here recommended to be used in parallel because the resulting estimates can be very sensitive to the choice of a parametric or nonparametric method. Consequently, the importance of a proper use of distribution selection methods should not be underestimated. Statistical tests, graphical exploration and expert knowledge can help in identifying the appropriate distribution. Both parametric and nonparametric methods have their advantages and disadvantages and their use depends on the expert's opinion, the problem formulation, the goals and the sample size. Parametric methods assume that the data come from a fixed form underlying distribution. This assumption enables them to work with smaller sample sizes. Non-parametric methods rely on the data points themselves. This makes them less vulnerable to deviations from certain distribution assumptions but more vulnerable to deviations in the data points. When the sample size is small (e.g. below 10), preference could be given to parametric methods whereas when the sample size is very large (e.g. above 50), preference could be given to nonparametric methods. For the intermediate sample sizes, either parametric or nonparametric techniques or both could be used.

Box 4: Criteria to consider when selecting a parametric distribution

First, graphical methods can provide valuable insights and generally could be used in conjunction with exploratory data analysis. They reveal important characteristics of a data set, including skewness (asymmetry), number of peaks (multi-modality), behaviour in the tails, and data outliers. Examples of graphical methods are frequency distributions (i.e. histograms), dot plots, line plots for discrete distributions, box-and-whisker plots and scatter plots. In a QQ-plot, observed values of a single numeric variable are plotted against the values that would be obtained if the sample were from a normal distribution. If the sample is from a normal distribution, points will cluster around a straight line. Here, the line is plotted through the first and third quartile of the data. The QQ-plot also depends on plotting positions and those are calculated according to Hazen.

Second, expert judgement refers to inferential opinion of a specialist or group of specialists within an area of their expertise. When there is uncertainty or variability associated with an input variable, such as

a data gap, expert judgement may be appropriate for obtaining distributions. For example, for the following related parameters in a metal risk assessment, distributions could be selected such that are given:

- The pH can not be smaller than 0 and larger than 14.
- An essential metal may have a minimum threshold level in its SSD below which ecotoxicity effects do not occur (except deficiency effects).
- Concentrations should always be positive numbers.

Distributions based on expert judgement can also serve as Bayesian priors in a decision-analytic framework. The distributions and Bayesian priors can be modified as new empirical data become available.

Third, goodness-of-fit tests exam how well (or poorly) a sample of data can be described by a hypothesized probability distribution for the population. Goodness-of-fit tests are formal statistical tests of the hypothesis that the data represent an independent sample from an assumed distribution. These tests involve a comparison between the actual data and the theoretical distribution under consideration. For the purpose of deriving the tail percentiles (e.g. HC_5 estimate), preference is given to the outcome of e.g. the Andersen-Darling test because it places more emphasis on tail values. However, goodness-of-fit tests have low statistical power and often provide acceptable fits to multiple distributions. Thus, goodness-of-fit tests are better used to reject poorly fitting distributions than for ranking good fits. For small n, goodness-of-fit tests will often fail to reject many of the hypothesized probability distributions.

Once all parameters in the exposure and effects assessment are characterized as either point estimates or probability distributions, it should be checked whether there are correlations between these parameters. This is important for subsequent Monte Carlo simulation.

4.2.2 Uncertainty propagation through exposure, effect or risk models

Once the most important input parameters are identified and their uncertainty and variability is quantified, the uncertainty and variability of the input parameters should be propagated to the ECD, SSD and finally risk quotient. There are a variety of ways to propagate information about variability or uncertainty through a model. A good reference with an extensive overview of techniques is Cullen & Frey (1999). The most common technique is Monte Carlo analysis (more information in box 5).

Box 5: Monte Carlo analysis

In Monte Carlo analysis, random samples of model input parameters are selected according to their respective assigned probability distributions. Once the samples from each input distribution are selected, the set of samples is entered into the deterministic model (e.g. E-USES, risk quotient,…). The model is then solved as it one would do for any deterministic analysis. The model results are stored and the process is repeated until the specified number of model iterations is completed. Instead of obtaining a discrete number for model outputs (as in a deterministic simulation) a set of output samples is obtained from which the output distribution can be characterized (Cullen and Frey 1999). In this way difficulties to estimate model input parameters and taking into account the inherent uncertainty and variability in specific processes are overcome.

Vose's (1996) 'cardinal rule of risk analysis modelling' is "Every iteration of a risk analysis model must be a scenario that could physically occur". If e.g. a high river flow is selected ad random, then a low temperature will be more likely than a large one if the river flow is highly negatively correlated with the temperature. Therefore, one of the restrictions that must be placed on the model is to recognize inter-dependencies between its uncertain components. It is possible to simulate jointly distributed random variables in which correlations may exist. Guidance can be found in Vose (1996).

Box 6 : first order versus second order Monte Carlo analysis

A first order or one-dimensional Monte Carlo simulation can only propagate variability or uncertainty, but not both at the same time without having difficulties with interpreting the output. It is therefore recommended that for this, among other propagation techniques, a second order or two-dimensional or embedded Monte Carlo simulation could be applied (Cullen & Frey, 1999). It consists simply in two Monte Carlo loops nested one inside the other. The inner one deals with the variability of the input variables, while the outer one deals with uncertainty. For each shot of a (uncertain) parameter value in the outer loop a whole distribution is created in the inner loop based only on variability (see Figures below). In this way changes in variability-dependent frequency distributions under the influence of parameter uncertainty can be quantified.

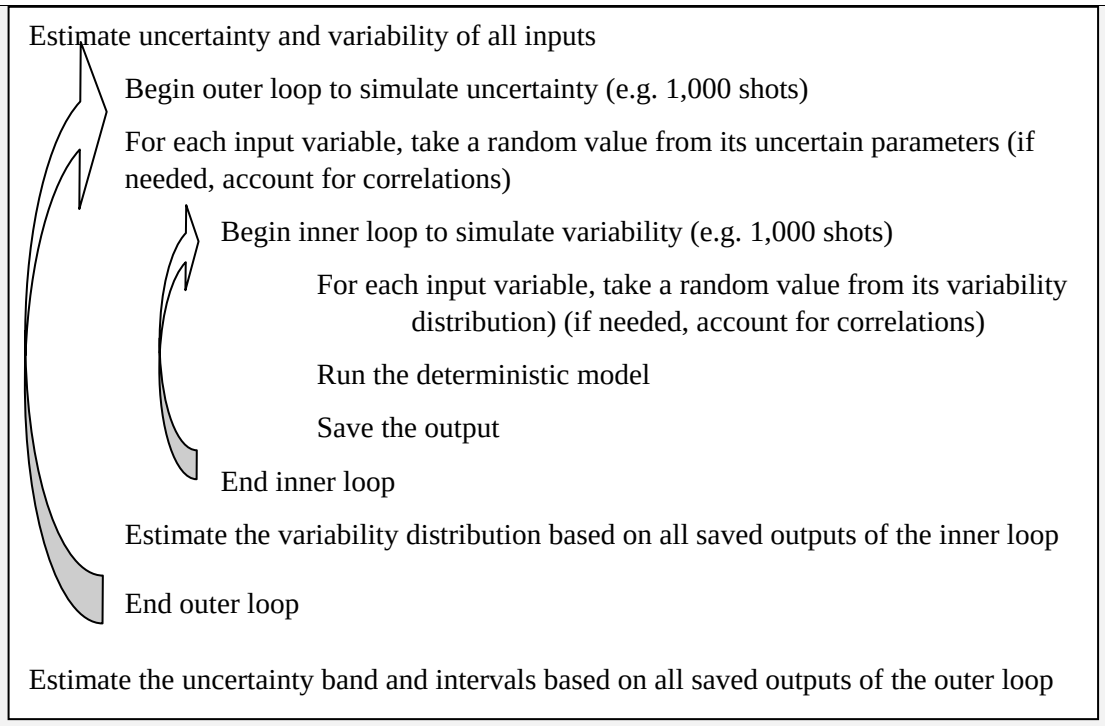

Figure 5: Simulation algorithm of a second order Monte Carlo simulation

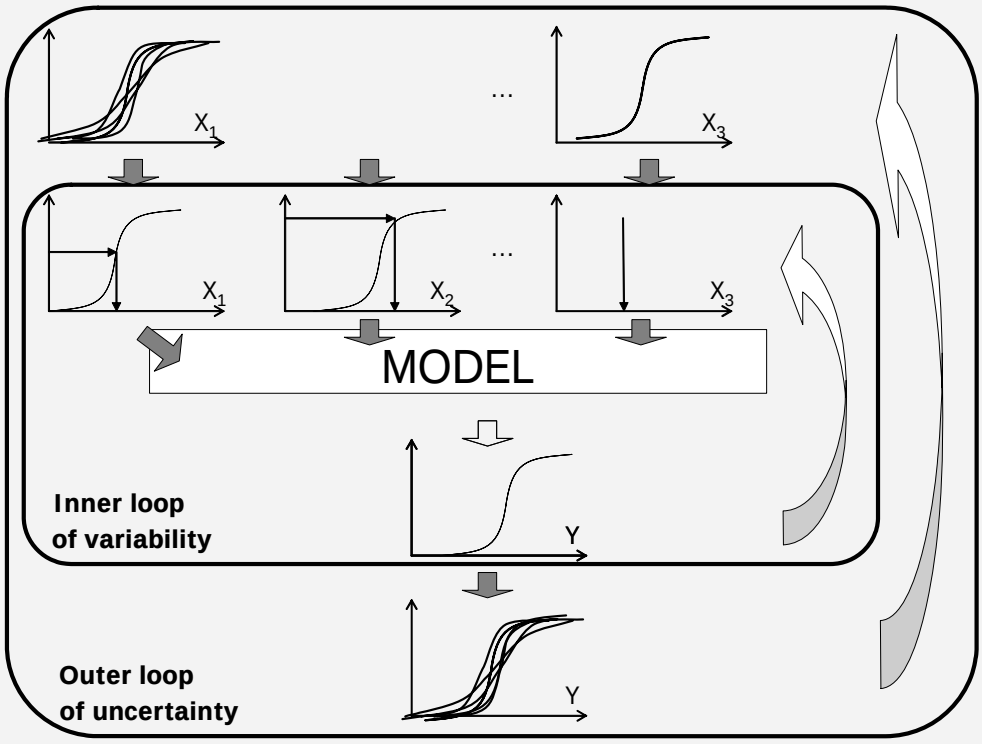

Figure 6: Graphical simulation algorithm of a second order Monte Carlo simulation (X_1 is uncertain and variable, X_2 is mainly variable, X_3 is mainly uncertain)

The final result of an uncertainty analysis on the risk assessment of a data-rich metal is an Exposure Concentration Distribution (ECD) or a Species Sensitivity Distribution (SSD) with an uncertainty or

confidence band in respectively the exposure and effect assessment. An example is given in Figure 7. The uncertainty is visualized as 90% confidence bands around the ECD and SSD. Note that in this example, only the sampling uncertainty of measured exposure concentrations is considered and that this uncertainty is very small since a lot of data were available.

The uncertainty bands transparently communicate the level of uncertainty to the risk managers. It quantifies how reliable the ECD and the SSD estimations are. This allows risk managers to better use the risk assessment results in a weight-of-evidence approach.

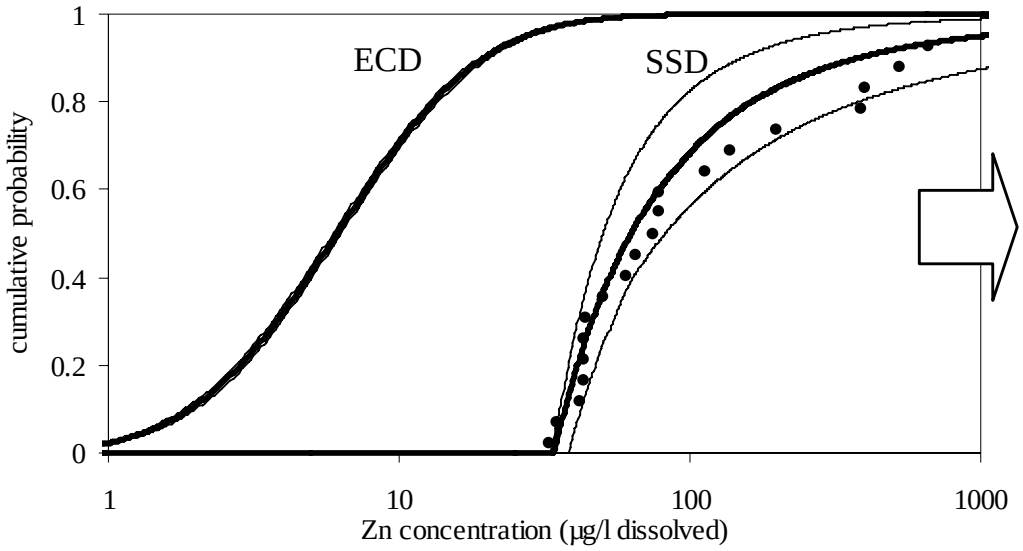

Figure 7: Example of uncertainty analysis for Zn in the Dutch surface waters (from: Van Sprang et al., 2004)

Not all sources of variability and uncertainty can be quantified. Model uncertainty and decision rule uncertainty for example are difficult to quantify and propagate through the assessment. These remaining sources of variability and uncertainty could in this case be added to the assessment on a (semi-)qualitative basis to the extent possible.

The uncertainty can subsequently be propagated to a probabilistic risk. More information on probabilistic risk characterization can be found in the respective fact sheet on risk characterization (MERAG fact sheet 1).

REFERENCES

Aldenberg T. & Jaworska J. S. 2000. Estimation of the hazardous concentration and fraction affected for normally distributed species sensitivity distributions. *Ecotoxicology and Environmental Safety*, 46, 1-18.

Aldenberg T., Jaworska J. S. & Traas T. P. 2002. Normal species sensitivity distributions and probabilistic ecological risk assessment. Pages 49-102 *in* L. Posthuma, G. W. Suter II & T. P. Traas, editors. Species sensitivity distributions in ecotoxicology. Lewis publishers, London, UK.

Burmaster D. E. & Wilson A. M. 1996. An introduction to second-order random variables in human health risk assessments. *Human and Ecological Risk Assessment*, 2(4), 892-919.

Cardwell R. D., Brancato M. S., Toll J., DeForest D. & Tear L. 1999. Aquatic ecological risks posed by tributyltin in United States surface waters: pre-1989 to 1996 data. *Environmental Toxicology and Chemistry*, 18(3), 567-577.

Cullen A. C. & Frey H. C. 1999. Probabilistic techniques in exposure assessment. A handbook for dealing with variability and uncertainty in models and inputs. Plenum, New York. 335.

Davison A. C. & Hinkley D. V. 1997. Bootstrap methods and their application. Cambridge University Press. 582.

ECOFRAM. 1999. ECOFRAM Aquatic and Terrestrial Final Draft Reports. US Environmental Protection Agency.

Efron B. & Tibshirani R. J. 1993. An introduction to the bootstrap. Chapman & Hall, New York.

EPA. 1997. Guiding Principles for Monte Carlo Analysis. EPA/630/R-97/001, Environmental protection Agency.

EPA. 1999. Report of the workshop on selecting input distributions for probabilistic assessments. EPA/630/R-98/004, U.S. Environmental Protection Agency, Washington, DC.

EPA. 2001. Risk assessment guidance for superfund (RAGS): Volume III - Part A. Process for conducting probabilistic risk assessment. US Environmental Protection Agency.

Giesy J. P., Solomon K., Coats J. R., Dixon K. R., Giddings J. M. & Kenaga E. E. 1999. Chlorpyrifos, ecological risk assessment in North American aquatic environments. *Reviews in Environmental Contamination and Toxicology*, 160, 1-129.

Giesy J. P., Solomon K., Coats J. R., Dixon K. R., Giddings J. M. & Kenaga E. E. 1999. Chlorpyrifos, ecological risk assessment in North American aquatic environments. *Reviews in Environmental Contamination and Toxicology*, 160, 1-129.

Gumbel E. J. 1958. Statistics of extremes. Wiley Classics Library, John Wiley and Sons, New York, USA.

Hart A. 2001. Probabilistic risk assessment for pesticides in Europe: implementation & research needs. *European workshop on probabilistic risk assessment for the environmental impacts of plant protection products*. The Netherlands.

Hazen A. 1914. Storage to be provided in impounding reservoirs for municipal water supply. *Transactions of the American society of civil engineers*, 77, 1539-1640.

Helsel D.R. and Hirsch R.M. 2002. Statistical methods in water resources. Techniques of Water-Resources Investigations of the United States Geological Survey. Book 4, Hydrologic Analysis and Interpretation. Chapter A3. 510 pages.

Iman R. L. & Conover W. J. 1982. A distribution-free approach to inducing rank order correlation among input variables. *Commun Statist-Simula Computa*, 11(3), 311-334.

NRC. 1983. Risk assessment in the federal government: managing the process. National Academy Press, Washington DC.

Jager T. 1998. Uncertainty analysis of EUSES: interviews with representatives from Member States and industry. RIVM report 679102047. National Institute of Public Health and the Enviroment, Bilthoven, The Netherlands.

Jager T., den Hollander H.A., Janssen G.B., van der Poel P., Rikken M.G.J., Vermeire T.G. 2000. Probabilistic risk assessment for new and existing chemicals: sample calculations. RIVM report 679102049. National Institute of Public Health and the Enviroment, Bilthoven, The Netherlands.

Jager T., Rikken M.G.J., van der Poel P. 1997. Uncertainty analysis of EUSES: improving risk managment by probabilistic risk assessment. RIVM report 679102039. National Institute of Public Health and the Enviroment, Bilthoven, The Netherlands.

Jager T., Vermeire T. G., Rikken M. G. J. & van der Poel P. 2001. Opportunities for a probabilistic risk assessment of chemicals in the European Union. *Chemosphere*, 43, 257-264.

Posthuma L., Suter II G. W. & Traas T. P., editors. 2002. Species sensitivity distributions in ecotoxicology. Lewis publishers. 587 Pages.

Rai S. N., Krewski D. & Bartlett S. 1996. A general framework for the analysis of uncertainty and variability in risk assessment. *Human and Ecological Risk Assessment*, 2(4), 972-989.

Saltelli, A., Chan, K., Scott, E. M. 2000. Sensitivity Analysis. Wiley Series in Probability and Statistics. ISBN 0-471-99892-3, 486 Pages.

Solomon K., Baker D. B., Richards R. P., Dixon K. R., Klaine S. J., La Point T. W., Kendall R. J., Weisskopf C. P., Giddings J. M., Giesy J. P., Hall L. W. & Williams W. M. 1996. Ecological risk assessment of atrazine in North American surface waters. *Environmental Toxicology and Chemistry*, 15(4), 31-76.

RIP. 2005. Preliminary Technical Guidance Document. Part C: Reference preliminary technical guidance document. REACH Implementation Project 3.2-1B. http://ecb.jrc.it/

Solomon K., Giesy J. P. & Jones P. 2000. Probabilistic risk assessment of agrochemicals in the environment. *Crop Protection*, 19, 649-655.

Solomon K. R. 1996. Overview of recent developments in ecotoxicological risk assessment. *Risk Analysis*, 16, 627-633.

Solomon K. R. & Takacs P. 2002. Probabilistic risk assessment using species sensitivity distributions. Pages 285-314 *in* L. Posthuma, G. W. Suter II & T. P. Traas, editors. Species sensitivity distributions in ecotoxicology. Lewis publishers.

The Cadmus Group I. 1996. Aquatic ecological risk assessment software user's manual, version 2.0. RP91-AER1, Water Environment Research Foundation, Alexandria, VA, USA.

van Asselt MBA. 2000. Perspectives on Uncertainty and Risk: The PRIMA approach to decision support, Kluwer Academic Publishers, Dordrecht, The Netherlands.

van der Sluijs JP, Risbey JS, Kloprogge P, Ravetz JR, Funtowicz SO, Corral Quintana S, Pereira AG, De Marchi B, Peterson AC, Janssen PHM, Hoppe R, Huijs SWF. 2003. RIVM/MNP guidance for uncertainty assessment and communication: Detailed guidance. RIVM/MNP (Guidance for uncertainty assessment and communication series, Vol 3). Report nr. NWS-E-2003-163. Utrecht University and RIVM, Utrecht/Bilthoven, The Netherlands. ISBN 90-393-3536-2.

Van Sprang P., Verdonck F.A.M., Vanrolleghem P.A., Vangheluwe M., Janssen C. 2004. Probabilistic environmental risk assessment of zinc in Dutch surface waters. Environmental Toxicology & Chemistry, 23 (12), 2993-3002.

Verdonck F.A.M., Van Sprang P., Vanrolleghem P.A. 2005. Uncertainty and precaution in European environmental risk assessment. Water Science & Technology, 52(6), 227-234.

Verdonck F.A.M., Aldenberg T., Jaworska J.S. & Vanrolleghem P.A. 2003. Limitations of current risk characterization methods in probabilistic environmental risk assessment. Environmental Toxicology and Chemistry, 22 (9), 2209-2213.

Vose D. 1996. Quantitative risk analysis. A guide to Monte Carlo simulation modeling. John Wiley & sons. 317.

Walker WE, Harremoës P, Rotmans J, van der Sluijs JP, van Asselt MBA, Janssen P, Krayer von Krauss MP. 2003. Defining uncertainty. A conceptual basis for uncertainty management in model-based decision support. Integrated Assessment, 4(1), 5-17.

Warren-Hicks W. J. & Moore D. R. J. 1995. Uncertainty analysis in ecological risk assessment. SETAC, Pensacola, US. 275.

Warren-Hicks W. J., Parkhust B. R. & Butcher J. B. 2002. Methodology for aquatic ecological risk assessment. Pages 345-382 *in* L. Posthuma, G. W. Suter II & T. P. Traas, editors. Species sensitivity distributions in ecotoxicology. Lewis publishers.