

# EUFRAM

Concerted action to develop a European Framework for probabilistic risk assessment of the environmental impacts of pesticides<sup>1</sup>

Work Package 4

## REVISED PRELIMINARY PAPER ON METHODS OF UNCERTAINTY ANALYSIS

January 2004

Peter Chapman and Melissa Reed, Syngenta:

With input from: Andy Hart, CSL; Tom Aldenberg, RIVM; Keith Soloman, CT UoG; Jose Tarazona, INIA; Matthies Liess, UFZ; Pamela Byrne, PCS, DAFRD.

## CONTENTS

1	Summary.....	2
2	Introduction.....	2
3	Problem formulation.....	4
4	General issues connected with fitting distributions .....	4
5	Distributions of measures of toxicity .....	7
6	Distribution of exposure.....	7
7	Combining exposure and effect distributions.....	8
8	Mathematical tools for probabilistic risk assessment.....	12
9	References.....	27
	APPENDIX 1: Species Sensitivity Distributions: History and Background to Methodology.....	28
	APPENDIX 2: Species Sensitivity Distributions: Some Thoughts on Sources of Variation.....	29
	APPENDIX 3: Overview of selected approaches for dealing with uncertainty in risk assessment.....	34

---

<sup>1</sup> EUFRAM is coordinated by CSL ([www.csl.gov.uk](http://www.csl.gov.uk)) and supported by European Commission's 5<sup>th</sup> Framework Programme ([www.cordis.lu](http://www.cordis.lu)), contract number QLK5 - CT 2002 01346. Further information and news about EUFRAM is provided at [www.eufram.com](http://www.eufram.com).

# 1 Summary

This paper discusses tools and methodology used in probabilistic risk assessment.

A key issue in any assessment is being clear about the problem we wish to address so we give guidance on problem formulation.

Probabilistic risk assessment necessarily involves fitting probability distributions so we provide a lengthy discussion on many of the issues. The precise course of action to be taken (when fitting a distribution) depends entirely on the specific problem, but also requires judgement, and often depends upon the availability of suitable software. So, whilst precise recommendations cannot be made for specific situations, general guidance is provided that should meet most people's needs.

A successful risk assessment involves combining a measure of toxicity with a measure of exposure. A successful probabilistic assessment therefore involves fitting distributions to both toxicity and exposure. The measure of toxicity we choose will depend upon the problem formulation, but in many cases it will take the form of a species sensitivity distribution (SSD). A concise history of methods that have been developed for estimating SSDs is presented in an appendix. Exposure distributions can be measured directly in the field or projected from models. Use of models requires an in depth knowledge of both the fate of toxicants in the environment and the behaviour of species exposed to these toxicants. Several hypothetical examples are presented to illustrate the point.

Projecting exposure distributions via models involves (1) determining which distributions are appropriate for individual components in the model and (2) making use of methods that allow these distributions to be combined into a distribution of exposure. A description of several useful methods is presented, together with an explanation of when each would be used. An example of each method is presented in the context of the daily dietary dose exposure model. The list of methods presented is as follows:

- Worst case analysis.
- Interval analysis.
- Fuzzy arithmetic.
- First-order (non-hierarchical or 1-D) Monte Carlo.
- Second-order (hierarchical or 2-D) Monte Carlo.
- Probability bounds (or P-bounds).

Finally, some thoughts are presented in an appendix on sources of variation in SSDs.

## 2 Introduction

A draft version of this paper was first produced prior to the first full meeting of the EUFRAM project in May 2003 and has been revised in the light of comments made both before and during that meeting. It discusses and reviews methodology for use in uncertainty analysis and recommends methods that can help us move towards a

harmonised approach within Europe. In particular it highlights methods that should be evaluated during the EUFRAM case studies (Work Package 8 or WP8).

Deterministic risk assessment in pesticide regulation involves combining a measure of toxicity with a measure of exposure in order to compute a measure of risk. In Europe a Toxicity Exposure Ratio (TER) is computed. In an early tier 1 assessment the measure of toxicity will be the smallest NOEC (or LD<sub>50</sub>) from a number of species and the measure of exposure will be a conservative estimate of exposure computed from a model. A ratio that takes a high value, i.e. greater than 5 or 10, indicates relative safety, whilst a low value indicates a relative risk. If a compound fails at tier 1 the assessment is refined via one or more steps that may include additional studies or additional modelling.

In probabilistic risk assessment (PRA) we seek to refine the assessment at all tiers, including tier 1, by recognising that toxicity varies between species and that exposure varies in time and place. This involves fitting frequency distributions to measures of both toxicity and exposure and then combining the distributions in appropriate ways in order to create a distribution of risk. The precise approach to combining distributions will vary and depend upon the problem being addressed. This paper therefore merely draws attention to appropriate methods in the PRA toolbox. Different approaches to applying the methods will be explored in WP8.

Practitioners of PRA draw a distinction between uncertainty and variability. This can be very confusing to the newcomer and takes some getting used to. The distinction is best grasped via an example. Suppose we transform a number of LD<sub>50</sub>s onto a log scale and then estimate the sample mean and sample standard deviation of a normal distribution. As part of our PRA we could use Monte Carlo methods to repeatedly sample from a normal distribution with parameters as estimated. The PRA practitioner would argue that we were taking account of variability, but not of uncertainty, because the sample mean and the sample standard deviation are merely estimates of the corresponding population values, not the actual values. In order to take account of uncertainty we would need to use a two-stage sampling process: first we would need to use Monte Carlo to sample a mean and a standard deviation from appropriate distributions, then using these estimates we would need to use Monte Carlo again to sample from a distribution of LD<sub>50</sub>s. In PRA variability refers to natural variability between items that cannot be changed, whilst uncertainty refers to lack of knowledge that can be improved upon. So, in the example already given the lack of precision in the estimate of the mean of the log(LD<sub>50</sub>s) can be reduced by taking a bigger sample of LD<sub>50</sub>s. In contrast, the variation between the LD<sub>50</sub>s cannot be reduced.

Section 4 discusses problem formulation. Section 5 provides general guidance on fitting distributions. Sections 6 and 7 discuss distributions for measures of toxicity and exposure respectively. Section 8 discusses ways of combining distributions of toxicity and exposure in order to produce a distribution of risk. Finally Section 9 presents a number of mathematical tools that have been found useful in PRA.

### 3 Problem formulation

The main principles of US EPA (1998) are applicable and a good starting point for EUFRAM. Key points include:

- Ensure the assessment endpoint is appropriate for the management goal and expressed in probabilistic terms.
- Develop a conceptual model that represents the mechanisms of exposure and effects and links available data to the assessment endpoint.
- Systematically examine the model to identify significant sources of uncertainty and variability, and dependencies.
- Write the model as formal equations.
- Check that the model is constructed at appropriate levels of temporal, spatial and biological scale.
- Check that any averaging or aggregation is done at appropriate levels of each scale.
- Decide on appropriate methods for propagating uncertainty and variability through the model and for incorporating dependencies.

### 4 General issues connected with fitting distributions

PRA naturally involves fitting probability distributions, so we begin by discussing general issues connected with fitting distributions.

Let us suppose that we have a sample of individuals, and that we are satisfied that this sample is representative (i.e. unbiased) of some population (i.e. collection of individuals) of interest to us. For example, the population could be the entire collection of LD<sub>50</sub>s of a specified group of avian species, or it could be the entire collection of exposures relevant to some risk assessment. Suppose we wish to fit a frequency distribution and then estimate features of interest that we will use in a risk assessment. Furthermore, we wish to recommend a harmonised approach to doing this. We have here a fairly straightforward problem in applied statistical inference, but with many options open to us, and many decisions to take. A list of issues needing to be resolved, in no particular order, might include:

- Which collection of probability distributions should we try fitting to observations?
- Should we transform data prior to fitting frequency distributions?
- Which methods should we use for fitting distributions – maximum-likelihood, method of moments, least squares?
- Should we use empirical approaches based on probability plots for fitting distributions? If so, which precise methods should we use?
- Should non-parametric approaches to fitting distributions be used – if so, which ones?
- Should we use goodness of fit tests for deciding which distributions are appropriate? If so, which ones?

- Should we consider using Bayesian approaches to fitting distributions (see section 6.2)? If so, should we use “pure” prior distributions that are based on expert judgement, or empirical prior distributions that are based on the historic database?
- Should we always use the same distribution for a given collection of species, or should we identify the best fitting distribution with every new set of data?
- What is the minimum number of species needed (more generally, how many individuals per test, how many tests per species, how many species)?
- From a fitted distribution of toxicity measures, which percentile(s) should we use as estimates of safe dose?
- Should we estimate percentiles or more conservative lower bounds on percentiles, for example as proposed by (Aldenberg & Slob, 1993)?
- How should we cope with censored observations?
- Should we refine estimates by using information from the historic database in a non-Bayesian way? If so, how?
- Should we weight fitting of distributions of toxicity measures as proposed by (Mineau et al., 1996).
- What are the implications of using an incorrect distribution on any specific occasion?
- Should we specify the level of precision required in estimates derived from probability distributions?
- When data clearly exhibit a mixture of two or more distributions, should we fit a single mixed distribution or partition the data into groups and fit separate distributions to each group?

Some of these questions have straightforward answers, some require judgement, some depend upon the precise specification of the problem, and some depend upon the availability of software tools. Some require further research. The remainder of this sub-section attempts to provide general guidance as far as is possible.

Transformation of data prior to fitting probability distributions will often be appropriate. For example, both measures of toxicity and measures of exposure will often need to be transformed onto a log scale prior to fitting. If in doubt fit the distribution both before and after transformation and decide which is best from the lack of fit test.

A straightforward, and perfectly acceptable, method of fitting a normal distribution involves fitting a straight-line regression model to the data. First the sample of numbers is ranked: the smallest value is given the value 1 and the largest value is given the number N. Then all the ranks are divided by (N+1) and converted to the probit scale (Collett; 1994). Finally, the probit transformed ranks are regressed against either the sample values or transformed sample values. The estimate of the mean and all percentiles can easily be obtained from the fitted line and the variance of the distribution is a function of the slope of the fitted line.

Maximum-likelihood is to be preferred over the fairly crude approach outlined above. With maximum-likelihood an initial guess is made for the values of parameters of a distribution and, using these values, the probability of obtaining the data is computed. This probability is known as the likelihood. A numerical optimisation

technique is then used to find the values of the parameters that maximise the likelihood. The method of least squares is identical to maximum-likelihood for normal data and should therefore only be used for fitting normal distributions (although it is possible to find references that suggest least squares is superior to maximum likelihood in all situations (Wold; 2000)). Many computer programs, e.g. SAS (2000) and Crystal Ball 2000, have routines for fitting a variety of distributions.

Following maximum-likelihood a goodness of fit test should be used to determine whether a specific distribution is appropriate for the data. A statistically significant result implies that the distribution is not appropriate. However the lack of significance does not prove a distribution to be correct. In practice, when several distributions are fitted to a set of data it is common to find several that are non significant, implying that any one of them could be consistent with the data. Resolving this situation requires judgement. One approach is to use the most straightforward of the candidate distributions, which is likely to have the fewest number of parameters. Another approach is to choose the distribution which gives the smallest p-value in the significance test.

In theory there is no limit to the number of distributions that we could consider. In practice we are unlikely to need to try more than a handful. The distributions found in commonly used computer programs, such as SAS (2000) and Crystal Ball 2000, are likely to be all that is needed.

Whether or not censored observations should be included in the fitting of distributions is very much down to the precise specification of the problem. There is no technical reason for not including censored observations when using maximum likelihood. An illustration for the normal distribution can be found in Johnson et al. (1994).

In some situations historic data can be used to improve the process of fitting distributions. This is particularly useful when the number of available data points is very small, making it difficult to decide upon a suitable distribution in the first place, and leading to very poor estimates of parameters. Bayesian methods are particularly suitable for incorporating historic data. With the Bayesian approach one has to determine the form of distribution for historic data and then specify a prior distribution for each parameter. The priors are then combined with the data to create posterior distributions for each parameter. For a good practical guide to Bayesian approaches see Gelman (1995).

Bayesian methods can also be used when there is no historic data. A non-informative prior (e.g. a uniform distribution) then have to be assigned to the parameters of the distribution. Because of lack of research on this approach, it is not possible to say whether it offers any advantages.

Non-Bayesian approaches to using historic data have also been proposed. An example is the method proposed by Luttick and Aldenburg (1997) for species sensitivity data (see Appendix 1). Such methods are perfectly acceptable.

On occasions it is clear that our sample comes from a mixture of two different distributions. This is easily seen as two intersecting straight lines in the straight-line

approach. There are many possible causes of this, but for species sensitivity the cause is often two different collections of species from different taxa. Our course of action depends upon the problem we wish to address. Often we will wish to allocate all sample values to one of two (or more) groups and fit a different distribution to each group. On occasion, however, we may wish to fit a single distribution to all groups.

Once fitted, a frequency distribution is used as part of the risk assessment procedure. Depending upon the problem we are trying to solve, this may involve estimating percentiles. For a comprehensive guide to wide range of continuous frequency distributions see Johnson et.al. (1995; 1995).

## **5 Distributions of measures of toxicity**

When we think of distributions of measures of toxicity we usually think of distributions of LD<sub>50</sub>s or distributions of NOECs. This is because there is so much published research into SSDs, but it doesn't need to be the case. Toxicity can refer to any measure that is particularly relevant to the problem at hand, for example, the response to a specific dose of a pesticide measured at multiple times and locations could form the basis of a distribution of toxicity provided it meets the requirements of the problem. Whatever measure is fixed upon, Section 3 provides advice on fitting distributions and estimating percentiles.

Because distributions of toxicity are so often represented by SSDs, there is a large, and growing, literature. Appendix 1 is a brief overview of the history of the subject and presents the main findings. Anyone wishing to fit SSDs needs refer to advice in Section 3 and be aware of the following:

- Fitting distributions by linear regression on transformed data (see Section 3.0) is a commonly used and very reasonable approach.
- Maximum likelihood approaches are better but require a greater level of skill and experience in order to apply judgement.
- The work of Aldenberg and Slob (1993) who explain how to estimate percentiles and confidence intervals around.
- The work of Luttick and Aldenberg (1997) who propose a method of using historic data to improve fitting when the sample is very small. This proposal has been well received and is recommended.
- The work of Mineau and Collins, who propose scaling factors to improve interspecies extrapolation.

## **6 Distribution of exposure**

The nature of exposure depends upon the specific context of the problem being tackled. Consider the following scenarios for example:

- Aquatic species living in water will be exposed to concentrations in water derived from spray drift, run off, or possibly leaching. If the chemical is water soluble and persistent the exposure will decrease slowly in still water over a

relatively long period. Alternatively, if the chemical has a short half-life, or quickly binds to sediment, exposure may only be for a very short period. In flowing water the exposure is likely to be very short.

- Aquatic species living in sediment are exposed to chemicals that enter water and then bind to the sediment. If the chemical is relatively persistent the exposure may decrease slowly over an extended period.
- Soil living terrestrial species, such as earthworms, will be exposed to concentrations derived from applications to crops. The depth through which exposure takes place will depend upon a number of things, such as whether or not the chemical is water soluble, whether it binds to a component of the soil, and whether or not it leaches. A narrower depth will usually correspond to a higher concentration. The concentration will decline slowly over many days if the chemical is relatively persistent, or disappear quickly if it is not. Depending on its lifestyle, and the depth at which it lives, a species may be exposed constantly or intermittently.
- Terrestrial species living on crops or in margins, or on the soil surface in crop or margin will be exposed from a number of sources – such as direct spray, by eating contaminated plant material, or by eating contaminated prey.
- Terrestrial species living outside a treated crop may venture into the crop to eat treated seed, treated vegetation, or contaminated vegetation. Their exposure will depend upon the proportion of contaminated food in their diet.

Concentrations of chemicals in soil, water, sediment, plants etc. can be obtained from direct measurement or from models. Furthermore, from the examples above it can be appreciated that estimating exposure requires an in depth knowledge of the fate of the chemical in the environment combined with a good understanding of the behaviour of species at risk.

Consider the situation in which exposure is a function of several entities, some or all of which are random variables that follow a probability distribution. First we have to decide which of the entities are constants and which are variables. We then have to decide which distribution is appropriate for each variable and determine parameters of each distribution (preferably estimating them from data). We have to derive the distribution of exposure by combining distributions and constants by some method or other (e.g. Monte Carlo). Finally we have to consider whether there are uncertainties that we need to take account of.

Section 7.0 gives a number of examples of deriving exposure distributions, all of them based on the formula for daily dietary dose.

## **7 Combining exposure and effect distributions**

In some (probably rare) circumstances we will be able to assume mathematical forms for the distribution of both sensitivity and exposure and to estimate parameters for both distributions. We can then combine exposure and sensitivity distributions mathematically and derive expressions for risk. (Aldenberg et al., 2002) provide a comprehensive account of the mathematical approach for the case in which both exposure and sensitivity follow a normal distribution.

Several graphical methods of presenting both distributions have been proposed and have found to be useful. These are often referred to as joint probability curves.

Figure 1 gives examples of two probability density functions (PDFs), one for exposure and one for sensitivity, both of which express relative frequency or relative probability. For the exposure distribution the area under the curve to the left of  $X$  represents the probability that the concentration of a toxicant in an environment selected at random will be less than  $X$ . Expressed as a percentage, this represents the percentage of environments in which the concentration of toxicant is less than  $X$ . For the sensitivity distribution the area under the curve to the left of  $X$  represents the probability that a species selected at random will have a toxicity measure (LD50, NOEC, etc.) less than  $X$ . Expressed as a percentage, this represents the percentage of species for which the toxicity measure is less than  $X$ .

**Figure 1.** Probability Density Functions (PDF's)

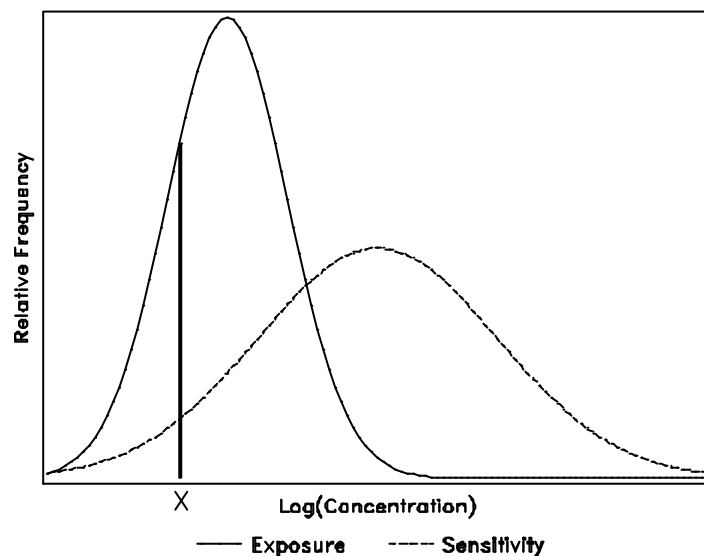
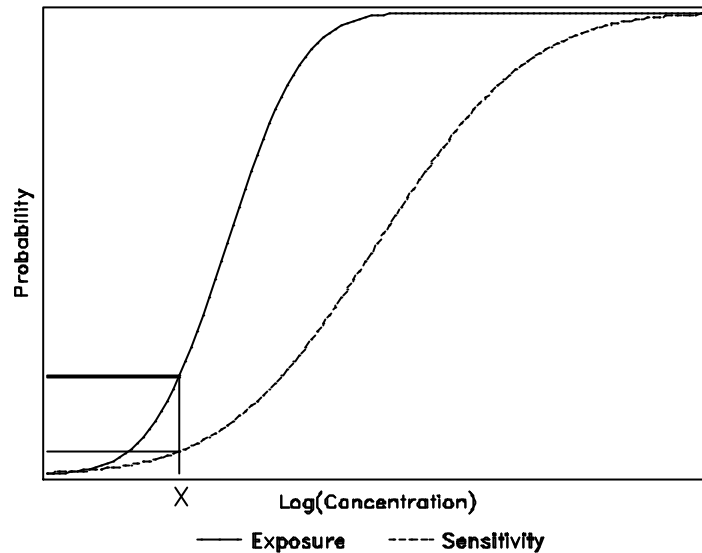


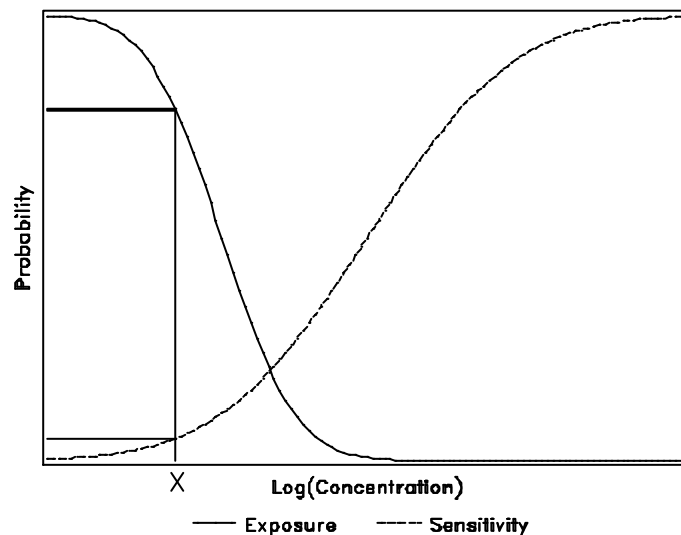
Figure 2 gives the cumulative distribution functions that correspond to the density functions in Figure 1. The cumulative functions are obtained by the mathematical process of integration.. The y-axis is now represents probability, so the areas under the curves to the left of  $X$  in Figure 1 become values on the y-axis in Figure 2.

**Figure 2.** Cumulative Distribution Functions (CDF's)



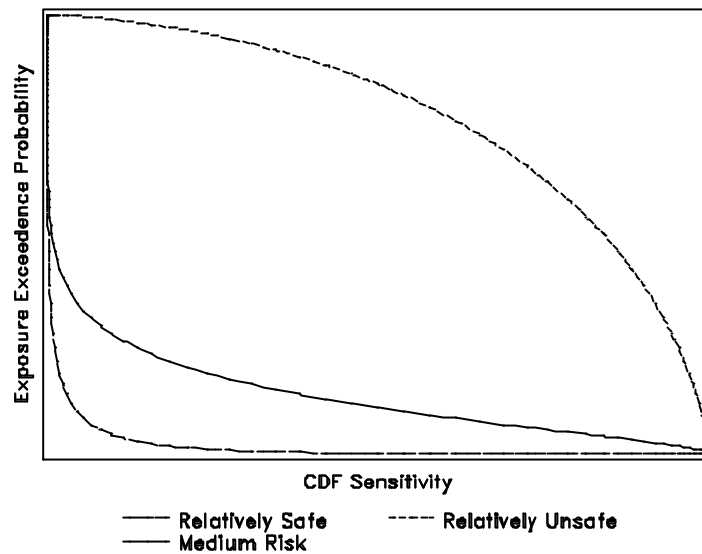
In Figure 3 the CDF for exposure is replaced by its inverse (i.e. 1 - the CDF). This is called the exceedence probability function (EXF) and, for each value of X on the x-axis, it gives the probability that the environmental concentration of the toxicant will be greater than X. Therefore if, for example, we select the concentration, X, that corresponds to 5% of species being affected (the 5<sup>th</sup> percentile on the pdf for sensitivity) then the EXF for the same value of X gives the probability that X will be exceeded, and hence that more than 5% of species will be affected.

**Figure 3.** Cumulative Distribution (Sensitivity) and Exceedence Probability (Exposure) Functions



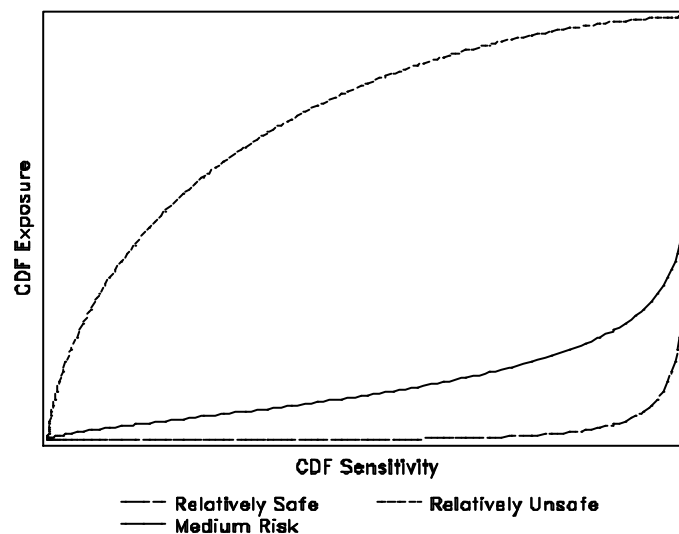
In Figure 4 the exceedence probability function (EXF) is plotted against the CDF for sensitivity. Graphs of this type are often presented in reports of PRAs and enable a direct read out of the probability that P% of species will be affected. Two extreme hypothetical cases are presented: in the graph: one in which the toxicant is perceived to be relatively safe, the other in which it is felt to be very unsafe.

**Figure 4.** Exceedence Probability Curve



Finally, in Figure 5 the CDF for sensitivity is plotted against the CDF for exposure. The interpretation of this graph is similar to the interpretation of Figure 4 and the hypothetical safe and unsafe curves are again illustrated.

**Figure 5.** Joint CDF Curves



Final point: If density functions can be expressed in mathematical form, then it is possible to compute the expected percentage of species affected by averaging over environments. Thus risk can be expressed as a single number rather than as a graph.

In some circumstances it may not be possible, or desirable, to express the distribution of both exposure and sensitivity in mathematical terms. An example might be when the distribution of exposure is obtained as an output from a model, and might be an empirical distribution that can be expressed graphically as a histogram. A Monte-Carlo procedure could be adopted whereby an exposure is

sampled at random and then paired with a sensitivity that is also sampled at random (taking account of any correlation between exposure and sensitivity). Empirical versions of the joint probability curves (Figures 1 to 5) can then be created and relevant measures of risk can be computed.

So far in this section we have discussed combining toxicity and exposure after distributions have been fitted to each. An alternative approach would be to combine the two before fitting distributions. So, for example, we could sample single values from the distributions of both toxicity and exposure, compute a TER, then carry on sampling in order to create a distribution of TERs. All of the methods outline in Section 7.0 below can be used with either approach.

## 8 Mathematical tools for probabilistic risk assessment

### 9.1 Introduction

The purpose of this section is to describe a number of tools that can be used for assessing variability and uncertainty. The precise list of tools covered is as follows:

- Worst case analysis.
- Interval analysis.
- Fuzzy arithmetic.
- First-order (non-hierarchical or 1-D) Monte Carlo.
- Second-order (hierarchical or 2-D) Monte Carlo.
- Probability bounds (or P-bounds).

The description of each tool is presented under five headings:

- A brief description of the method.
- A brief discussion of when one would use it.
- An example.
- A list of advantages.
- A list of disadvantages.

All of the examples are based on a hypothetical case in which we wish to estimate the exposure to a pesticide A, applied at 200g/ha, of a hypothetical insectivorous bird living in an orchard. The daily dietary dose (*DDD*) is computed from the following equation:

$$DDD = \frac{FIR}{BW} * C * PD * PT ,$$

in which C is the concentration on food, FIR is the food intake rate, BW is body weight of the bird, PT is the proportion of time spent in field, and PD is the proportion of insects in diet.

The data used in the examples are summarised in Table 1: maximum and minimum are used in worst case, interval and fuzzy methods, the best guess is used in fuzzy risk assessment, and the distribution type is used in the Monte Carlo and P-bound

approaches. Additional data is presented in the sections covering Monte Carlo and P-bounds. The food intake rate is negatively correlated with the body weight, but all other variables are assumed to be independent.

**Table 1.** Data for hypothetical insectivorous bird exposed to pesticide A applied at 200 g/ha.

Variable	Minimum	Maximum	Best Guess	Distribution	Mean	Std. Deviation/ Shape
C	2 µg/g	20 µg/g	6	Lognormal	6.97 µg/g	3.27 µg/g
FIR	0.4 g/day	1.2 g/day	0.7 g/day	Normal	0.8 g/day	0.155 g/day
BW	10 g	14 g	12	Normal	12 g	0.776 g
PT	0.11 (11%)	0.61	0.3	Lognormal	0.32	2.31
PD	0.7 (70%)	1	0.9	Normal	0.86	0.05

The worst-case analysis, interval analysis, fuzzy arithmetic, and probability bounds analysis were carried out using Risk Calc software (reference). The Monte Carlo analyses were carried out using Crystal Ball 2000.

Throughout the rest of this section the word variable is used to refer to the inputs to the daily dietary dose model – i.e. C, FIR, BW, PT and PD. The word parameter refers to the parameters of the distribution of a variable. For example, the mean and standard deviation of a normal distribution are parameters. We also distinguish between population values for parameters, which are fixed, but unknown, and estimates or sample values for parameters, which are known but follow a distribution. When we refer to the output from the model we are referring to the daily dietary dose. Depending on the method, the output can be a single value, or a distribution, or a set of bounds surrounding a distribution.

## 9.2 Worst Case Analysis

### *Description*

This method leads to a computation of the worst-case exposure by combining maxima and minima for each variable as is appropriate. The example below should make everything clear.

### *When and why one would use it?*

This is essentially a quick, and easy to use, screening tool. An examination of the inputs to the equations, and a sensitivity analysis, can identify major sources of

uncertainty that lead to a conservative outcome and highlight needs for additional work.

### **Example**

Daily dietary dose is calculated from the equation, using data from Table 1. The maximum values for C, FIR, PT and PD and the minimum values for BW lead to worst-case daily intake of 20.496 µg / day.

### **Advantages**

The method:

- Only requires a small amount of data.
- Is easy and quick to use.
- Is very conservative.
- Helps us identify areas where additional work should be focussed.

### **Disadvantages**

- Can be too conservative to be of practical value.

## **9.3 Interval analysis**

### **Description**

In interval analysis we use the maximum and minimum values for variables but make no assumptions about probability distributions. Since the input to the DDD exposure model is a range, the output is also a range. To illustrate, the sum of  $[x_1, x_2]$  and  $[y_1, y_2]$  (where both are ranges) is found by adding the two smallest numbers and the two largest to give the range  $[x_1+y_1, x_2+y_2]$ . The minimum for subtraction ( $[x_1, x_2] - [y_1, y_2]$ ) is found by subtracting the top of the range from the bottom ( $x_1 - y_2$ ) and the maximum by subtracting the bottom from the top ( $x_2 - y_1$ ). For example, addition of  $[3, 7]$  and  $[-4, 2]$  gives  $[-1, 9]$  and subtraction ( $[3, 7] - [-4, 2]$ ) gives  $[1, 11]$ . Rules also exist for multiplication, division, and other operators.

### **When and why you would use it?**

Interval analysis is useful when we can reasonably assume maximum and minimum values but have no knowledge of the relative likelihood of values between the extremes. Provided maxima and minima are accurate, the method is guaranteed to give an output range that encloses the true value, but tells us nothing about the relative likelihood of different values within the range.

### **Example**

The following ranges are all taken from Table 1:

C =  $[2, 20]$  µg / g

FIR =  $[0.4, 1.2]$  g / day

BW =  $[10, 14]$  g

$$PT = [0.11, 0.61]$$
$$PD = [0.7, 1],$$

The resulting interval for daily dietary dose is [0.044, 20.496]  $\mu\text{g} / \text{day}$ , which means that the output dose can be anywhere in the range 0.044 to 20.496  $\mu\text{g} / \text{day}$ . The maximum is identical to the result for the worst-case analysis.

### **Advantages**

- The method can handle any type of uncertainty.
- It is guaranteed to include the true value for the output variable provided the true values of all input variables lie within the ranges defined for them.
- It is simple and easy to explain
- It can be very conservative.

### **Disadvantages**

- Complex computations with multiple inputs can lead to very wide output intervals that provide little information.
- The method is often too conservative
- We learn nothing about the relative likelihood of values within the output intervals.

## **9.4 Fuzzy Arithmetic**

### **Description**

Fuzzy numbers are similar to intervals except that in addition to the maximum and minimum values the best guess is included for each of the input variables. For example, suppose the value of A is known to be between 3 and 6, so could be represented by the interval [3,6], but we also know that the most likely value is 5. We could represent this knowledge as the fuzzy number [3,5,6]. For the interval all values between the maximum and minimum are equally likely, but for the fuzzy arithmetic the best guess provided is the most likely value.

The output from a computation using fuzzy numbers is a minimum and maximum (the same as for interval analysis) plus a best estimate.

### **When and why one would use it?**

In order to use fuzzy numbers one needs to know the maxima and minima for each input variable and have a good idea as to the most likely value. As with interval analysis the output value is guaranteed to enclose the true value providing the input values are accurate, but also provides information about the relative likelihood of intermediate values.

### **Example**

In addition to the minimum and maximum values the best guess is used (see Table 1). The input values for the dietary daily dose are [minimum, best guess, maximum] are given below:

C = [2, 6, 20]  $\mu\text{g} / \text{g}$   
FIR = [0.4, 0.7, 1.2] g / day  
BW = [10, 12, 14] g  
PT = [0.11, 0.3, 0.61]  
PD = [0.7, 0.9, 1]

The output interval is 0.044 to 20.496  $\mu\text{g} / \text{day}$  (identical to that obtained from the interval analysis) with the most likely value at 1.134 (displayed as [0.044, 1.134, 20.49601]  $\mu\text{g} / \text{day}$ ).

### **Advantages**

- It has the same advantages as interval analysis.
- In addition it provides point estimates as extra information.

### **Disadvantages**

- It is not widely known or understood hence it is little used.
- If the central guess is inaccurate the method, by providing a single central point estimate of the output value, could mislead.

## **9.5 First-order (non-hierarchical) Monte Carlo**

### **Description**

Monte Carlo methods can be used when either:

1. We can express each of the input variables in terms of a mathematical probability distribution. This in turn requires us to have sufficient data in order to fit distributions.
2. We have a large set of historic data from which to sample.

The process is quite straightforward. First we sample at random a single value for each of the input variables using its respective distribution. Using these values we compute an output value from a model. We then repeat the whole process many times to give us a distribution of output values. How we use the output distribution will depend on the context, but will often involve presenting percentiles. Any correlations between input variables should be taken into account. A failure to do so can seriously distort the results. The number of repeat samples is also important – too few and the output distribution will not appear to be smooth. Experience and trial and error can help decide on the number of samples.

### **When and why one would use it?**

A Monte Carlo simulation can be used when

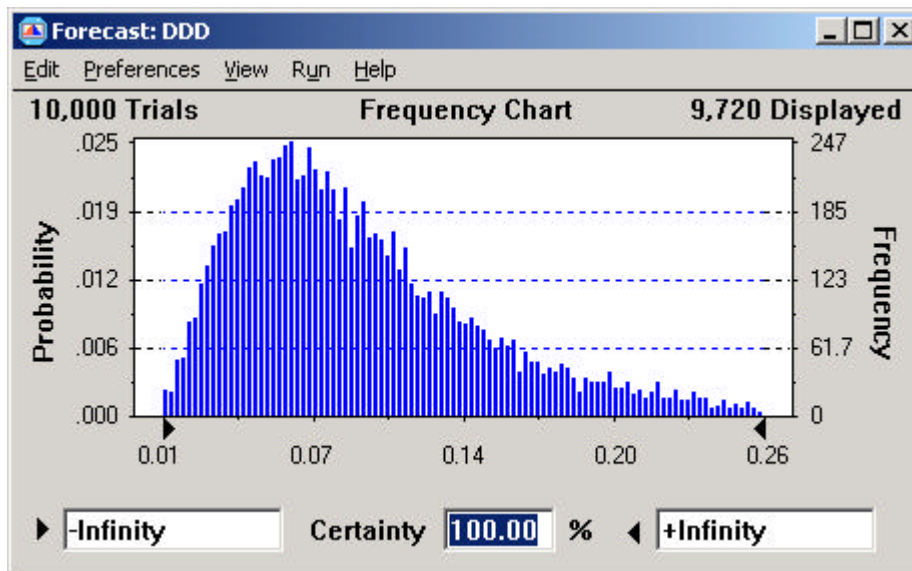
- There is sufficient data to define distributions of the input variables.
- We can quantify any correlations between input variables.

**Example**

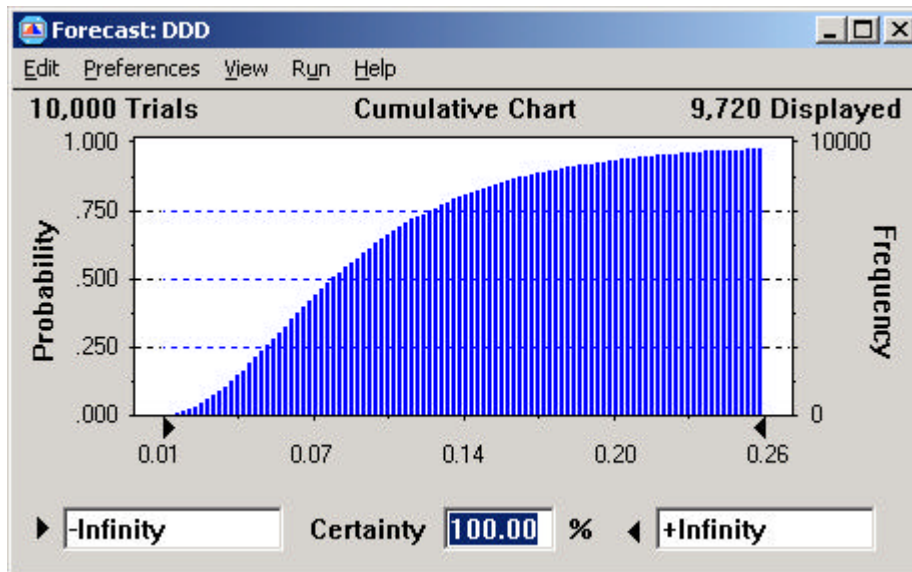
Distributions of input variables are taken from Table 1. The output distribution of daily dietary dose can be viewed in several different ways. Figure 6 shows the frequency chart, a histogram showing the number of times each result was produced. Figure 7 shows the cumulative frequency chart.

In addition to graphical output a large number of statistics can be computed, such as the mean (0.10  $\mu\text{g} / \text{day}$ ), standard deviation (0.06), median (0.08  $\mu\text{g} / \text{day}$ ), minimum (0.01  $\mu\text{g} / \text{day}$ ) and maximum (0.69  $\mu\text{g} / \text{day}$ ) and the percentiles.

**Figure 7.** First-order Monte Carlo: Frequency chart for daily dietary dose.



**Figure 8.** First order Monte Carlo: Cumulative frequency chart for daily dietary dose.



### Advantages

- The method is simple to use and understand and easy to explain.
- It can incorporate information about correlation between variables.
- It summarises the entire distribution of risk
- In addition to variability, it can incorporate sampling uncertainty about parameter estimates.

### Disadvantages

- The method requires large amounts of data.
- We need to make assumptions about the form of the distribution of each input variable. Often this can be difficult.
- It does not separate the effects of uncertainty (lack of knowledge) and variability (real world variation).
- It does not take account non-sampling types of uncertainty.

## 9.6 Second-order (hierarchical or 2D) Monte Carlo

### Description

Second order (or 2-D) Monte Carlo is an extension of first-order Monte Carlo. At the outset we classify each variable as either a source of uncertainty or a source of variability. We then proceed as follows. First we sample, at random, a single value from each variable that we have classified as uncertainty, whilst taking account, if appropriate, of any correlations between these variables. Using the values from this single sample we then perform a complete first-order Monte Carlo for all those variables that we classified as variability. We then repeat the process many times – i.e we take a single sample from the variables that we classified as uncertainty, followed by a complete first order Monte Carlo for those variables classified as variability. This has the effect of separating uncertainty from variability, as should become clear from the example below.

### ***When and why one would use it?***

Second order Monte Carlo analysis is useful when it is both possible and helpful to classify all variables as representing either variability or uncertainty. It is particularly helpful in enabling us to manage parameter or model uncertainty. In first-order Monte Carlo we repeatedly sample from input distributions to produce an output distribution, of exposure for example. The problem with this is that the original distributions from which we sample are estimated from data, or at least the parameters of the distribution are estimated and hence are subject to sampling error, but in first-order Monte Carlo we assume them to be known and fixed. In second-order Monte Carlo we can overcome this difficulty by setting up parameter uncertainty as a source of uncertainty in the outer loop of the simulation.

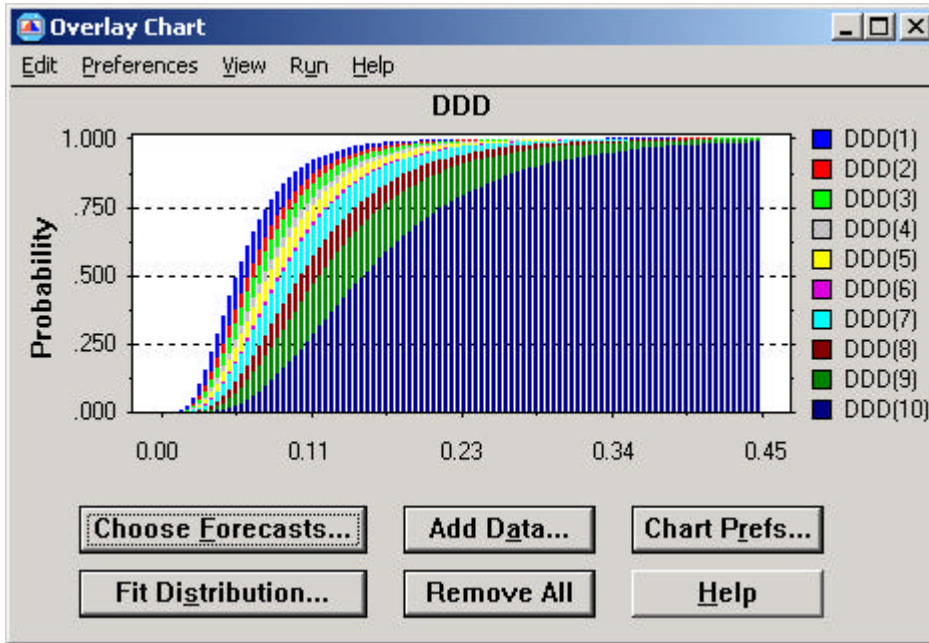
The easiest way to explain this is through the example of a normal distribution. Suppose a variable  $X$  follows a normal distribution with a population mean  $\mu$  and a population variance  $s^2$ . We use the sample mean  $\bar{x}$  and the sample variance  $s^2$  as estimates of the population mean and variance. We also know, from statistical theory, that the sample mean follows a normal distribution with a mean  $\mu$  and a variance  $s^2/n$ , where  $n$  is the size of the sample. In second order Monte Carlo we can classify lack of precision in the mean as uncertainty in the outer loop, and variability in  $X$  in the inner loop.

#### ***Example 1***

Distributions of input variables are again taken from Table 1. On this occasion, however, the proportion of time spent in field and the proportion of insects in the diet have been classified as uncertainty, with the remaining variables classified as variability. The outer loop (uncertainty) was sampled on ten occasions, whilst the inner loop (variability) was sampled 10,000 times for each sampling of the outer loop. Thus each sampling of the outer loop results in a complete distribution for the output from the inner loop. Figure 8 shows the cumulative distributions for all ten samples taken from the outer loop, and Table 2 gives the estimates corresponding to each run of the inner loop.

From study of the results it should become clear that second-order Monte Carlo allows us to construct a distribution of estimates. So, for example, not only can we now estimate a 5<sup>th</sup> percentile of exposure, we can also estimate the lower 5<sup>th</sup> percentile of the distribution of exposure.

**Figure 8.** 2-D Monte Carlo: Cumulative frequency distributions for daily dietary dose from ten runs of an outer loop.



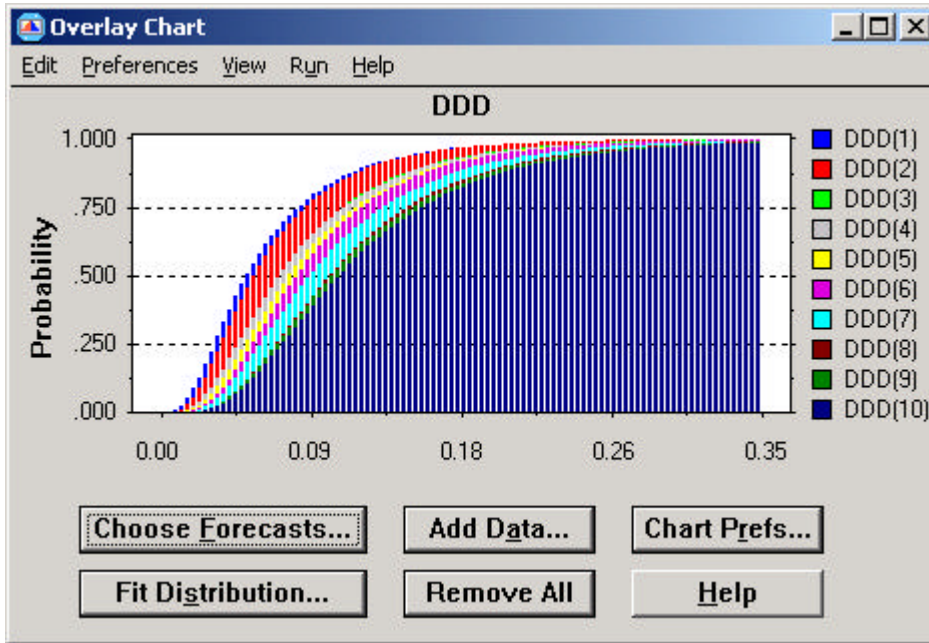
**Table 2.** 2-D Monte Carlo: Parameter estimates for ten runs of an outer loop.

Trial	Mean	St. Dev	Percentiles		
			5 <sup>th</sup>	50 <sup>th</sup>	95 <sup>th</sup>
1	0.054154	0.027065	0.021327	0.048252	0.107705
2	0.079248	0.040924	0.03121	0.070612	0.157614
3	0.090121	0.046539	0.035492	0.0803	0.179239
4	0.091405	0.047202	0.035998	0.081444	0.181793
5	0.096066	0.049609	0.037833	0.085597	0.191062
6	0.0971	0.050143	0.03824	0.086518	0.193118
7	0.109618	0.056607	0.04317	0.097672	0.218016
8	0.118623	0.061257	0.046717	0.105696	0.235925
9	0.121359	0.06267	0.047794	0.108133	0.241365
10	0.154705	0.07989	0.060927	0.137845	0.307687

### Example 2

In the first example the variation in three of the variables was assumed to be due to variability and in the other two to uncertainty. It is also possible to include both variability and uncertainty for a single variable. In this example the proportion of time spent in the field is still assumed to be follow a lognormal distribution, but there is uncertainty about the value of the mean. The mean of 0.25 is replaced by a normal distribution of mean 0.25 and standard deviation of 0.05. The cumulative frequency chart from this simulation is shown in Figure 9. Estimates of parameters describing the distributions are shown in Table 3.

**Figure 9.** 2-D Monte Carlo: Cumulative frequency distributions for daily dietary dose: ten runs of the outer loop with uncertainty and variability included for PT



**Table 3.** 2-D Monte Carlo: Parameter estimates for ten runs of an outer loop: PT classified as both variability and uncertainty.

Trial	Mean	St. Dev	Percentiles		
			5 <sup>th</sup>	50 <sup>th</sup>	95 <sup>th</sup>
1	0.084933	0.070596	0.026172	0.070596	0.191676
2	0.093935	0.078501	0.029433	0.078501	0.210215
3	0.094687	0.079401	0.029914	0.079401	0.210794
4	0.094918	0.079325	0.029739	0.079325	0.212439
5	0.0975	0.081147	0.030177	0.081147	0.219483
6	0.099033	0.083176	0.03144	0.083176	0.219388
7	0.103753	0.087319	0.033101	0.087319	0.229358
8	0.10784	0.091989	0.036057	0.091989	0.234859
9	0.133703	0.114755	0.045957	0.114755	0.287664
10	0.141628	0.122208	0.049718	0.122208	0.303212

### Advantages

- The method can separate out uncertainty and variability.
- It can handle model uncertainty in a limited way.

### Disadvantages

- Simulations can take a lot of computer power to run.
- Parameterisation can be difficult.
- Results are difficult to present and explain.
- It does not take account of uncertainty about distribution shape.

## 9.7 Probability bounds analysis

### Description

Suppose we know that a variable follows a normal distribution with a mean of between 19 and 21, and a standard deviation of between 1 and 3, but we know no more. Then probability bounds analysis (or P-bounds) allows us to put limits around a cumulative distribution for the variable. The interval between the bounds represents our state of ignorance about the distribution. Complete knowledge of the distribution gives bounds that are coincident, which implies zero uncertainty, although with variability from some underlying stochastic process.

Probability bounds can also be used when the form of the distribution is unknown. In fact, bounds in the output distribution can be created from whatever information is available. This may be sample data or information on a combination of minimum, maximum, mean, mode, median, percentile, variance, standard deviation and others. One can mix and match distributions, intervals, point values and data. For example, if we know that variable A has a normal distribution with a mean of between 7 and 8,5 and a standard deviation of 2.3 to 2.9 but we only know that B has a minimum of 0 and a maximum of 2 we may still proceed. It is also possible to nominate the dependencies between variables.

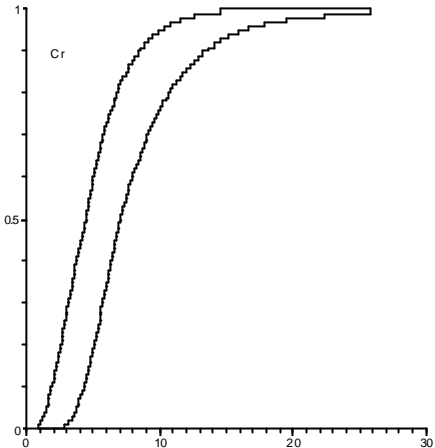
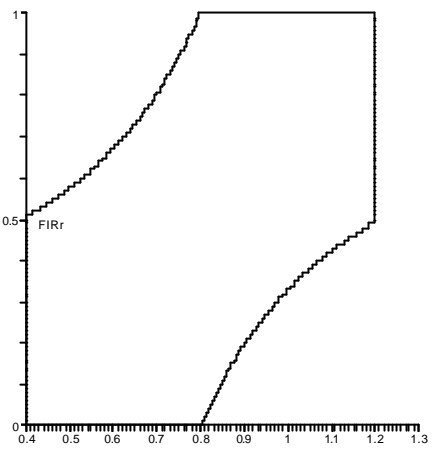
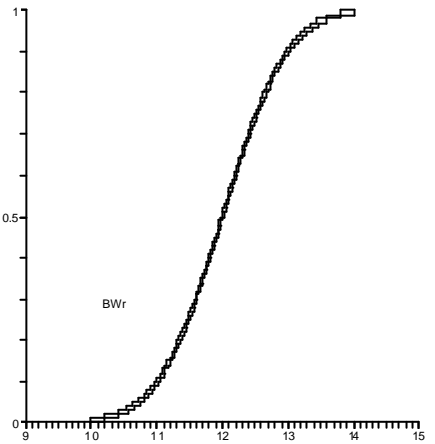
### *When and why you would use it?*

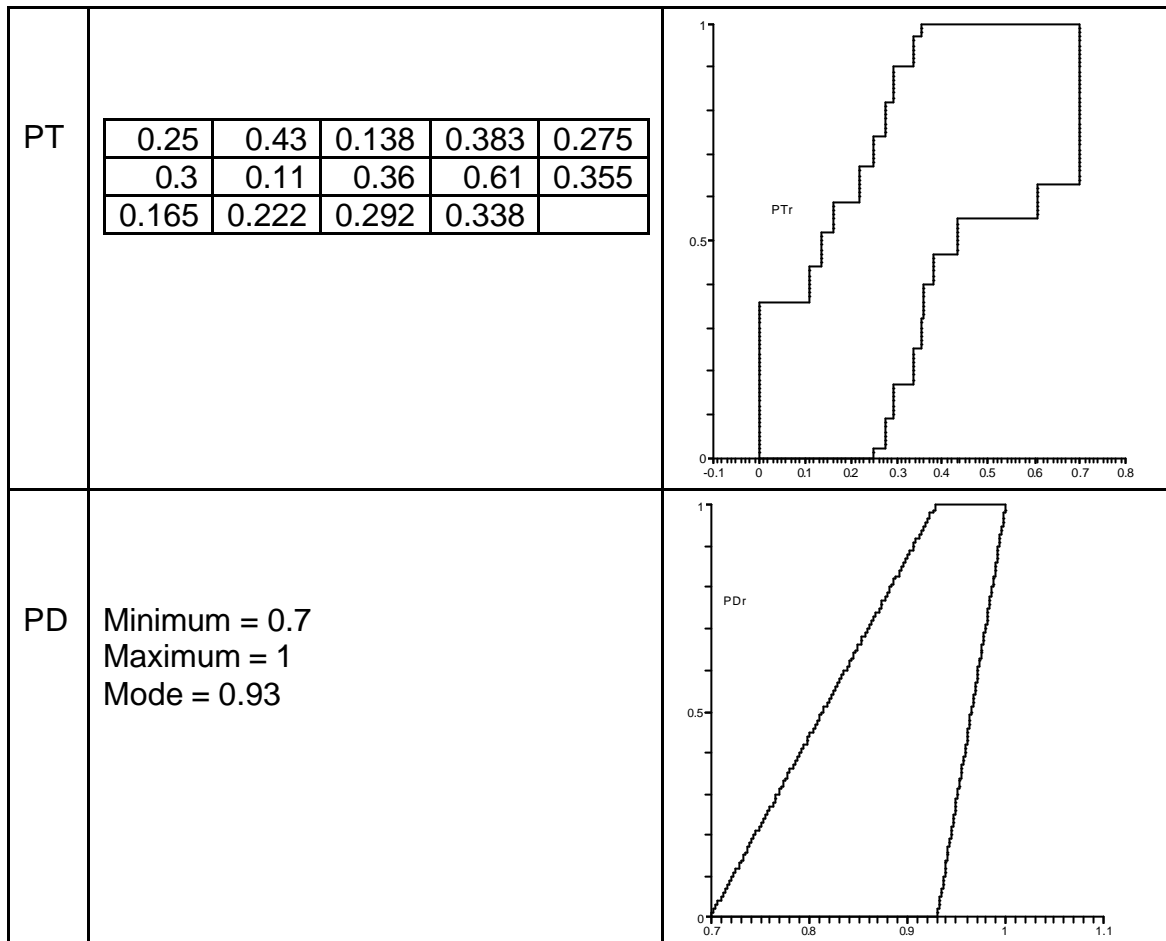
P-bounds can be used when we know that input variables follow distributions but there is a lack of knowledge, and hence a great deal of uncertainty.

### *Example*

The data used in this example is also taken from Table 1. The concentration is still assumed to be a lognormal distribution, but there is uncertainty about the parameters of the distribution. The mean is now assumed to be in the interval [5.5, 7.5] with a standard deviation in the interval [3, 4]. We no longer assume the food intake to be normally distributed, as previously, but here only the minimum (0.4), maximum (1.2) and mean (0.8) are used. We have a lot of data with which to estimate body weight so a normal distribution with a mean of 12  $\mu\text{g}$  and a standard deviation of 0.776  $\mu\text{g}$  seems OK. However, we introduce an interval to reflect the number of significant digits recorded (We can handle this in Risk Calc by enclosing the number in square brackets). We have fourteen values sampled from the distribution of PT, and in this example we have simply used the data without attempting to fit a distribution. For PD we have simply used the minimum, maximum and mode. Table 3, for each variable, describes in detail the assumptions we have made and illustrates the bounds on the distribution that these assumptions translate to.

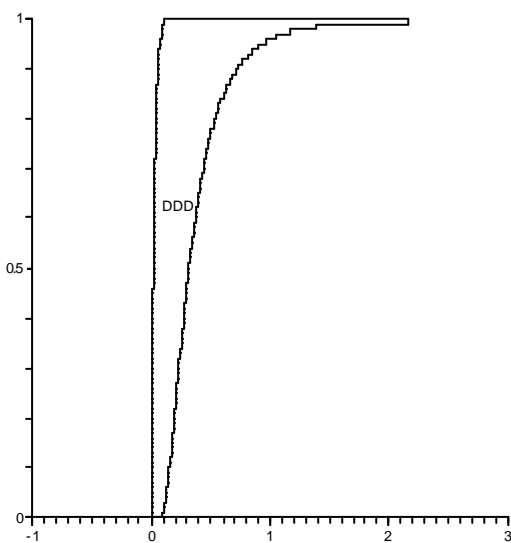
**Table 4.** Probability bounds analysis: Input distributions used in the example.

	Description	Graph
C	Lognormal distribution Mean = [5.5, 7.5] $\mu\text{g} / \text{kg}$ St Dev = [3, 4]	
FIR	Minimum = 0.4 Maximum = 1.2 Mean = 0.8	
BW	Normal distribution Mean = [12.00] St Dev = [0.776]	



The set of assumptions are combined to produce bounds on the output distribution of the daily dietary dose. As with the previous examples independence was assumed between all variables except food intake rate and body weight. Figure 10 shows the result from this analysis. The mean, range and variance are shown underneath.

**Figure 10.** The cumulative distribution for daily dietary dose calculated using P-bounds



Mean = [0.038,0.263]  
Range = [0,2.17379]  
Variance = [0,0.097]

### **Advantages**

- The method can account for different types of uncertainty – i.e. uncertainty about parameter values, distribution shapes, dependencies, and model form.
- Output distributions are presented in the form of bounds.
- Bounds get narrower, and hence uncertainty reduces, as we gain more knowledge about variability and uncertainty of input variables.

### **Disadvantages**

- Infinite tails on distributions of sampling uncertainty must be truncated.
- The method lacks a theoretical basis from which to interpret its handling of sampling uncertainty.

## **9.8 Summary of Examples**

Since we have based all the examples on the same set of data and the same model for computing daily dietary dose, it is instructive to compare results from all methods in the Table 7.

**Table 7.** A comparison of methods used in the examples in Section 9.

Method	Data Requirements	Result	DDD example
Worst Case Analysis	Worst case values for each parameter	A worst case estimate of risk	20.496 µg / day
Interval Analysis	Intervals giving the highest and lowest value for each parameter	An interval containing the true risk	0.044 - 20.496 µg / day
Fuzzy Risk Assessment	Intervals giving the highest and lowest value and a best guess for each parameter	An interval containing the true risk with an indication of the most likely value	0.044 - 20.496 µg / day with a most likely value of 1.134 µg / day
First order Monte Carlo	Distribution types with defining variables or a data set to fit a distribution to for each on input parameter	Frequency chart showing the results of the trails and statistics about the resulting distribution, for example mean, standard deviation and the percentiles from 0% to 100%	Mean = 0.10 µg / day Standard Deviation = 0.06 50 <sup>th</sup> Percentile = 0.08 µg / day 90 <sup>th</sup> Percentile = 0.17 µg / day
Second order Monte Carlo	Distribution types with defining values and knowledge on where the variation in due to uncertainty and where it is due to variability	A set of frequency distributions representing the uncertainty of the input parameters and a table showing parameters about each distribution (e.g. mean, standard deviation, 5 <sup>th</sup> to 95 <sup>th</sup> percentiles)	Mean from 0.084933 µg / day to 0.141628 µg / day 95 <sup>th</sup> percentile from 0.191676 µg / day to 0.303212 µg / day 50 <sup>th</sup> percentile from 0.070596 µg / day to 0.122208 µg / day
Probability Bounds Analysis	Distributions for the input variables with bounds defining the uncertainty about them, data to fit bounds to or statistics about the data for example minimum, maximum and mean. Also information about dependencies between variables	A boundary that the true distribution of risk falls within.	Mean from 0.038 µg / day to 0.263 µg / day Variance from 0 to 0.097

## 9 References

Aldenberg, T., Jaworska, J.S., & Traas, T.P., eds. (2002) Normal species Sensitivity Distributions and Probabilistic Ecological Risk Assessment. Lewis, CRC Press, Boca Raton.

Aldenberg, T. & Slob, W. (1993) Confidence Limits for Hazardous Concentrations Based on Logistically Distributed NOEC Toxicity Data. *Ecotoxicology and Environmental Safety*, 25, 48-63.

Forbes, T.L. & Forbes, V.E. (1993) A Critique of the Use of Distribution-Based Extrapolation models in ecotoxicology. *Functional Ecology*, 7, 249-254.

Forbes, V. & Calow, P. (2002) Species Sensitivity Distributions Revisited: a Critical Appraisal. *Human and Ecological Risk Assessment*, 8, 473-492.

Jago, R.H. & Newman, M.C. (1997) Bootstrap estimation of community NOEC values. *Ecotoxicology*, 6, 293-306.

Kooijman, S.A.L.M. (1987) A Safety Factor for LC50 Values Allowing for Differences in Sensitivity Among Species. *Water Research*, 21, 269-276.

Luttik, R. & Aldenberg, T. (1997) Extrapolation Factors for Small Samples of Pesticide Toxicity Data: Special Focus on LD50 Values for Birds and Mammals. *Environmental Toxicology and Chemistry*, 16, 1785-1788.

Mineau, P., Collins, B.T., & Baril, A. (1996) On the Use of Scaling Factors to Improve interspecies extrapolation of acute toxicity in birds. *Regulatory Toxicology and Pharmacology*, 24, 24-29.

van Straalen, N.M. & Denneman, C.A.J. (1989) Ecotoxicological Evaluation of Soil Quality Criteria. *Ecotoxicology and Environmental Safety*, 18, 241-251.

Wagner, C. & Løkke, H. (1991) Estimation of Ecotoxicological Protection Levels from NOEC Toxicity Data. *Water Research*, 25, 1237-1242.

## **APPENDIX 1: Species Sensitivity Distributions: History and Background to Methodology**

(Kooijman, 1987), (van Straalen & Denneman, 1989), (Wagner & Løkke, 1991) and (Aldenberg & Slob, 1993) are the four papers that established the theoretical foundation for fitting probability distributions to effects measured on a number of species. Their combined contribution to the subject was as follows:

- A frequency distribution of known parametric form, such as a log-normal or log-logistic function could be fitted to effects measured on a number of species - using maximum likelihood, or some other method.
- The effects they chose to use in their examples were NOECs, LD50's and such like, because these were readily available.
- Having fitted the distribution the lower p'th percentile, called HC<sub>p</sub>, could be estimated. This was proposed as an upper level for a safe exposure concentration that would protect (100-p) % of species.
- Because HC<sub>p</sub> is an estimate of a percentile, the true value of which is unknown, it will exceed the true value approximately 50% of the time. The lower X% confidence limit on the estimate of HC<sub>p</sub> was proposed as a more conservative "safe concentration" which would be larger than the true value on only (100-X)% of occasions. Typically authors specifically referred to the lower 5<sup>th</sup> percentile, HC<sub>5</sub>, and to the lower 95<sup>th</sup> confidence limit.

Whilst mathematically satisfying, some authors felt that there were real practical issues and difficulties that could invalidate the method. These views are best summarised by (Forbes & Forbes, 1993), who highlighted four assumptions that need to be satisfied if the methodology is to be considered valid:

- i. The distribution of species sensitivities in natural ecosystems should closely approximate the postulated theoretical distribution.
- ii. The sensitivities of species used in the laboratory should represent an unbiased sample from the sensitivity distribution of species in natural communities.
- iii. By protecting species composition, community function is also protected.
- iv. Interactions among species in communities or ecosystems can be ignored.

(Forbes & Forbes, 1993) concluded that these assumptions had not been rigorously tested and were probably not valid, implying in turn that the methodology was flawed.

Following on from these early papers a number of authors attempted either to address the assumptions or to develop the methodology in various ways. These are summarised below:

- (Jagoe & Newman, 1997) used bootstrap methods to estimate safety concentrations. They argued that this non-parametric approach overcomes the need to satisfy assumptions i. and ii. listed above. Whilst this is no doubt true for i., it is hard to see how a non-parametric method can overcome sampling bias.
- Newman et. al. (1999) conclude that no single type of distribution is appropriate with real data sets on more than 50% of occasions and that bootstrapping overcomes deficiencies connected with fitting mathematical distributions.

They also conclude that between 15 and 55 species are required in order to estimate the HC<sub>5</sub>, HC<sub>10</sub>, and HC<sub>20</sub> for a specific compound.

- (Luttik & Aldenberg, 1997) recommend that the methods developed by (Aldenberg & Slob, 1993) should be used. They proposed using the method without modification when toxicity measures are available for four or more species. For three or fewer species they recommended computing a pooled estimate of variance from the historic database. Whilst these recommendations were made specifically for avian species, they translate easily to other groups of species.
- (Mineau et al., 1996) recommend (again for avian species) that average weight of the species needs to be taken account of when fitting a distribution. This was based on an observation that the relationship between log (LD50) and log (weight) was a straight line but that the slope of the line was not 1.

The mathematical treatment for a large number of specific scenarios has been worked out. In this sense the topic of species sensitivity distributions can be said to be in an advanced state of development. For example, (Aldenberg et al., 2002) provide a comprehensive treatment for a normal distribution under the assumption that a collection of log effect concentrations is an unbiased sample from a field population of interest. In another sense, however, much less progress has been made in addressing concerns, all of which centre on whether assumptions are valid.

Finally, (Forbes & Calow, 2002) reiterate some of the issues raised in (Forbes & Forbes, 1993) and provide some guidance on how the SSD approach can be improved. In addition to the issues already mentioned above, they argue that:

- That the sample of species used should be relevant to the community of species about which we wish to make an assessment.
- That endpoints for input into SSDs should be more rigorously chosen, and should include endpoints that measure likely population level impacts.
- That more careful consideration should be given to the fraction of species we wish to protect and that the fraction should vary according to the total number of species at risk.

## **APPENDIX 2: Species Sensitivity Distributions: Some Thoughts on Sources of Variation**

### **A2.1 Introduction**

This appendix shows that different sources of variation and the testing structure we put in place to estimate SSDs affect both (a) the underlying distribution that we wish to estimate plus (b) the precision with which we estimate the distribution. A possible implication is that even with a random process for selecting species, laboratories and so forth, a minimal testing approach may result in a biased estimating procedure.

The fact that tested species, for example, are not selected at random from the pool of all species (about which we are interested) merely adds to the problem.

Before optimum procedures can be established for estimating SSDs, a complete understanding of (a) what sources of variation are important, (b) their structure relative to each other, (c) and the relative size of each variance component, is needed. This understanding will enable optimum strategy:

- Number of animals per test.
- Number of tests per species.
- Number of species.

Clearly gaining a good understanding of sources of variation requires a great deal of work and so is well beyond the scope of the Eufram project. Nevertheless, it does represent a worthwhile research project for consideration in future.

## A2.2 Components of Variation in the Context of Laboratory Testing.

In order to understand the nature and size of the variance of some outcome or statistic, statisticians often model the problem in terms of “components of variation”. To illustrate the method, a (very) simplified model of the variation observed in tolerance of individual animals is constructed here.

Suppose we select an organism at random from a population comprising different species, different suppliers of laboratory animals, different labs, hypothetical repeats of the same test at different times within a lab, and all of the organisms included in a single test. Then the organism is selected from a frequency distribution of the form:

$$Q(\mu: s_{sp}^2 + s_s^2 + s_{lab}^2 + s_{tlab}^2 + s_{atlab}^2),$$

Where

Q = frequency distribution of unspecified type, such as a normal or logistic - this is presented with two parameters, a mean and variance, but could be generalised to three or more parameters,

$\mu$  = mean tolerance for entire population – e.g. log(LD50),

$s_{sp}^2$  = variance between species,

$s_s^2$  = variance between suppliers,

$s_{lab}^2$  = variance between labs,

$s_{tlab}^2$  = variance between occasions within labs,

$s_{atlab}^2$  = variance between animals within occasions within labs.

Let us suppose that species, suppliers and laboratories are cross-classified, that occasion is nested within these, and that animal is nested within occasion. For SP species, S suppliers, L laboratories, T occasions, and A animals, the variance of the tolerance of an individual animal is:

$$S = s_{sp}^2 + s_s^2 + s_{lab}^2 + s_t^2 + s_a^2,$$

the mean value of the tolerance is:

$$= \mu,$$

and the variance of the mean tolerance is

$$= \{1/(SP*S*L)\}\{s_{sp}^2 + s_s^2 + s_{lab}^2 + (1/T)(s_t^2 + (1/A)s_a^2)\}.$$

The equivalent formulae for a single species, denoted by i, are:

$$s_s^2 + s_{lab}^2 + s_t^2 + s_a^2,$$

$$\mu_i = \mu + di,$$

$$\{1/(S*L)\}\{s_s^2 + s_{lab}^2 + (1/T)(s_t^2 + (1/A)s_a^2)\}.$$

The components of variance  $s_{sp}^2, s_s^2, s_{lab}^2, s_t^2$  and  $s_a^2$ , can be estimated by selecting at random a number of species, suppliers, labs, carrying out a number of tests within each lab, and measuring the tolerance of a number of individual animals selected at random. Methods of computing the estimates (sample values)  $s_{sp}^2, s_s^2, s_{lab}^2, s_t^2$  and  $s_a^2$  will not be discussed here.

Once the components of variance have been estimated we can use them to help design future studies.

### A2.3 Optimum Allocation of Resource

Suppose we wish to estimate, HC5, the lower 5'th percentile of the frequency distribution for a collection of species. Then,

$$HC5 = \mu - Ks = \mu - K*\sqrt{\{1/(S*L)\}\{s_{sp}^2 + s_s^2 + s_{lab}^2 + (1/T)(s_t^2 + (1/A)s_a^2)\}},$$

where K is a positive constant independent of the number of species. We estimate this by substituting sample values s for s, and m for  $\mu$ , i.e.,

$$est(HC5) = m - Ks = m - K*\sqrt{\{1/(S*L)\}\{s_{sp}^2 + s_s^2 + s_{lab}^2 + (1/T)(s_t^2 + (1/A)s_a^2)\}},$$

The expected value of this estimate depends upon S, L, T, and A as well as well as absolute values of  $s_{sp}^2, s_s^2, s_{lab}^2, s_t^2$  and  $s_a^2$ . Thus, increasing the number of suppliers and labs, tests per lab/supplier and animals per test all lead to larger values of HC5. Furthermore, for a given total number of animals, the expected value of the estimate depends on the allocation of animals and the relative values of  $s_{sp}^2, s_s^2, s_{lab}^2, s_t^2$  and  $s_a^2$ . For example, if  $s_t^2$  is much bigger than  $s_a^2$  and we reduce the number of animals per test whilst increasing the number of tests per lab/supplier/species (keeping total number of animals constant) then the expected value of HC5 will get larger.

In practice we usually have only one test per species so S, L and T are all equal to 1. The expected value of the estimate then becomes,

$$HC5 = \mu - Ks = \mu - K*\sqrt{[s_{sp}^2 + s_s^2 + s_{lab}^2 + s_t^2 + (1/A)s_a^2]},$$

The precision of the estimate of HC5 is determined by the standard deviations of the estimates m of  $\mu$ , and s of s, which are both functions of

$$s/SP = \{1/(SP*S*L)\}\{s_{sp}^2 + s_s^2 + s_{lab}^2 + (1/T)(s_t^2 + (1/A)s_a^2)\},$$

which is estimated by,

$$s/SP = \{1/(SP*S*L)\}\{s_{sp}^2 + s_s^2 + s_{lab}^2 + (1/T)(s_t^2 + (1/A)s_a^2)\}.$$

So, the precision of the estimate of EC5 depends not only on the number of species tested but also on the allocation of animals to tests. Depending on the relative values of  $s_{sp}^2, s_s^2, s_{lab}^2, s_t^2$  and  $s_a^2$ , it may be possible to improve precision (reduce standard error) of the estimate of HC5 by using fewer animals per test and increasing the number of tests per species.

## A2.4 Sampling Bias

The discussion in A1.2 focused on expected value of HC5. The expected value is equal to the mean (or average) value obtained from a hypothetical infinite number of repeat computations. However, in practice we will only compute HC5 once.

If, as in current practice, we use one test (implying only one supplier and lab) per species, then for a single computation of HC5 the expected value is:

$$HC5 = \mu - Ks = \mu + d_s + d_{lab} + d_t - K*\sqrt{[1/(S*L)\{s_{sp}^2 + 1s_a^2/A\}]}$$

Thus a single estimate of HC5 could be described as biased, the bias emerging because only one test is carried out at one lab and using animals from one supplier. These bias terms do not disappear if we increase the number of tests, labs, and suppliers, but they do get smaller, so overall bias reduces.

This argument suggests that we can encounter bias even if we use random processes for selecting lab, supplier, and for selecting the time of the test. Add to this argument the fact that the selection is not random, and that species themselves are not selected at random, we have to conclude that in the context of the lab we are using a biased process to estimate HC5 (for example).

Notes:

- The sources of variation discussed here – lab, supplier etc. – were selected for illustration purposes. Many other sources could be identified.

- The structure used here for the component of variance model was for illustration. It is never easy to decide upon the exact form for such a model and the arrangement of species, lab and supplier in a cross-classified form may be incorrect.
- Thus, setting up a model in practice requires a great deal of thought and discussion on (a) appropriate sources to include in a model and (b) an appropriate structure for the model.
- Throughout this document we have assumed that tolerance of an individual can be measured. Of course this is not true. All we can do is dose an individual and the result (death or survival) tells us whether the dose is below or above that individual tolerance. Making this false assumption does not invalidate the argument.

**APPENDIX 3: Overview of selected approaches for dealing with uncertainty in risk assessment.**

	<b>How?</b>	<b>Why?</b>	<b>Why not?</b>
<b>Worst case analysis</b>	<ul style="list-style-type: none"> <li>• estimate assuming the plausible extreme</li> <li>• compare with reference value</li> </ul>	<ul style="list-style-type: none"> <li>• account for uncertainty by being conservative</li> <li>• under ignorance, shift burden of proof</li> </ul>	<ul style="list-style-type: none"> <li>• level of conservatism unquantified and may be too low or too high</li> </ul>
<b>Interval analysis</b>	<ul style="list-style-type: none"> <li>• replace each point estimate with an interval (e.g. [1,2])</li> <li>• use interval arithmetic to combine the intervals</li> </ul>	<ul style="list-style-type: none"> <li>• natural for scientists and easy to explain to others</li> <li>• works no matter where uncertainty comes from</li> </ul>	<ul style="list-style-type: none"> <li>• paradoxical: can't give exact value but can give exact bounds</li> <li>• ranges can grow very quickly, giving very wide results</li> </ul>
<b>Monte Carlo simulation</b>	<ul style="list-style-type: none"> <li>• replace each point estimate with a probability distribution</li> <li>• repeatedly sample from each, tally answers in a histogram</li> </ul>	<ul style="list-style-type: none"> <li>• simple to implement</li> <li>• fairly simple to explain</li> <li>• summarizes entire distribution of risk</li> <li>• can use information about correlations between variables</li> <li>• user-friendly software on familiar platforms</li> </ul>	<ul style="list-style-type: none"> <li>• requires a lot of empirical information – or assumptions</li> <li>• assumptions can lead to non-protective conclusions</li> <li>• only appropriate if uncertainty is statistical</li> <li>• does not separate uncertainty and variability</li> </ul>
<b>Second-order Monte Carlo</b>	<ul style="list-style-type: none"> <li>• let parameters of input distributions be distributions too</li> <li>• nest Monte Carlo analyses</li> <li>• summarize with distribution of distributions, or condense into a single distribution</li> </ul>	<ul style="list-style-type: none"> <li>• acknowledges and accounts for uncertainty about distribution parameters</li> <li>• separates variability and uncertainty</li> <li>• can handle model uncertainty in a limited way</li> <li>• user-friendly software on familiar platforms</li> </ul>	<ul style="list-style-type: none"> <li>• can be daunting to specify inputs</li> <li>• requires data or assumptions about distribution shape and dependencies</li> <li>• results are cumbersome to interpret and explain</li> <li>• confounds frequentist and subjectivist interpretations of probability</li> </ul>

	How?	Why?	Why not?
<b>Robust Bayes (Bayesian sensitivity analysis)</b>	<ul style="list-style-type: none"> <li>• identify a class of distributions to represent uncertainty about the prior distribution</li> <li>• identify a class of functions to represent uncertainty about the likelihood function</li> <li>• form the class of posterior distributions by applying Bayes rule</li> </ul>	<ul style="list-style-type: none"> <li>• accounts for the analyst's doubts about required inputs</li> <li>• can be much cheaper and easier than insisting on precise inputs</li> <li>• consistent with robust statistics</li> <li>• expresses the reasonableness of the posterior</li> </ul>	<ul style="list-style-type: none"> <li>• can be computationally difficult</li> <li>• does not obey Bayesian "dogma of ideal precision"</li> </ul>
<b>Probability bounds analysis</b>	<ul style="list-style-type: none"> <li>• specify what you are sure about</li> <li>• establish bounds on probability distributions</li> <li>• pick dependencies (no assumption, independence, correlated, perfect, etc.)</li> </ul>	<ul style="list-style-type: none"> <li>• handles uncertainty about parameter values, distribution shapes, dependencies, and model form</li> <li>• puts rigorous bounds on Monte Carlo results</li> <li>• bounds get narrower with better empirical information</li> <li>• faithful to frequentist interpretation of probability</li> </ul>	<ul style="list-style-type: none"> <li>• displays must be cumulative</li> <li>• does not yield 2D probabilities</li> <li>• must truncate infinite tails</li> <li>• lacks theoretical basis for interpreting its treatment of sampling uncertainty</li> <li>• difficulties with repeated variables or complex dependencies</li> </ul>
<b>Range-moment propagation</b>	<ul style="list-style-type: none"> <li>• for each variable estimate range and moments {min, max, mean, variance}</li> <li>• (or it'll work with {one endpoint, mean, variance} or {mean, variance} or {one endpoint, mean} or just {min, max})</li> <li>• use formulas to propagate range and moments through binary operations (+, -, ×, etc.)</li> <li>• express results as bounds on the tail risks (it looks like a p-box)</li> </ul>	<ul style="list-style-type: none"> <li>• very modest (and non-daunting) data requirements</li> <li>• combines advantages from interval analysis and basics of probability theory</li> <li>• especially useful in making a screening assessment fully probabilistic</li> <li>• compatible with a probability bounds analysis</li> </ul>	<ul style="list-style-type: none"> <li>• cannot use further information about distribution and correlations (must resort to probability bounds)</li> <li>• may yield answers that are very wide</li> </ul>