Uncertainty analysis

Eric Smith

Volume 4, pp 2283–2297

in

Encyclopedia of Environmetrics
(ISBN 0471 899976)

Edited by

Abdel H. El-Shaarawi and Walter W. Piegorsch

Uncertainty analysis

In risk assessment, available information is collected and utilized to make decisions regarding the risk associated with a particular stressor such as a chemical, biological, or physical agent (*see* **Risk assessment, management and uncertainties**). Decisions in risk assessment are typically not crystal clear and hence there is uncertainty. Uncertainty analysis is the part of risk assessment that focuses on the uncertainties in the assessment. Important components of uncertainty analysis include qualitative analysis that identifies the uncertainties, quantitative analysis of the effects of the uncertainties on the decision process, and communication of the uncertainty. The analysis of the uncertainty depends on the problem. An analysis of uncertainty for a risk assessment of **global warming** would be quite different than that for **hazardous agents** such as polychlorinated biphenyls (PCBs). Differences result from differences in spatial and temporal scale, available data and information, models and objectives.

Uncertainty analysis, as applied in the ecological, medical, and general risk analysis literature, therefore covers a wide range of techniques and analyses. The analyses range from simple descriptive procedures to quantitative estimation of uncertainty, to more formal decision-based procedures. The analysis may be qualitative or quantitative, depending on the level of resolution required and the amount of information available. The assessment of uncertainty is also tied to the view of uncertainty from the scientist and risk manager.

Views of Uncertainty

A scientist's view of uncertainty often varies with field. A risk manager often will view uncertainty in terms of a decision process, evaluating the errors associated with decisions and their costs (*see* **Economics, environmental**). Uncertainty is viewed as a nuisance factor that clouds the decision process. A toxicologist might view uncertainty in terms of a model that relates risk to toxic stress (*see* **Toxicology, environmental**). Uncertainty is then associated with the models selected, the variation in organisms, and measurement error. A statistician might view uncertainty in terms of the manner in which data are collected and analyzed. Often the uncertainty is summarized in terms of bias, variance, and measures based on the statistical distribution. Many other views of uncertainty are possible including the chemist (uncertainty in measurement), the sociologist (uncertainty associated with communication of risk), and the citizen.

Risk is often viewed as the probability of an undesired or harmful event. The connection to probability is implied by the uncertainty in the occurrence of the event. The definition of uncertainty – as the lack of surety or certainty – is readily defined in a statistical or probabilistic context as the implication that uncertainty exists when the probability of an event occurring is not zero or one [11]. It is no surprise then that statistics and probability provide the main conceptual tools for uncertainty analysis (*see* **Risk assessment, quantitative**).

Although the uncertainty analysis problem is conceptually similar to statistical problems, in application the uncertainty analysis problem is often more complex than many statistical problems. There often are multiple populations, data are often not sampled to address a problem – rather, information on a problem is collected from different sources – and it may not be possible to estimate all parameters in the model. Epidemiologists and ecological risk analysts thus take a view of risk and uncertainty that is more general than the statistical view (*see* **Risk assessment, ecological**). The model that is used in complex risk assessments is itself complex, involving multiple sources of information and multiple models that may be on different scales. Researchers involved with these models often categorize uncertainty into three components: structural, parameter, and stochasticity. Structural uncertainty refers to uncertainty due to lack of knowledge about the correct model. Parameter uncertainty is associated with the uncertainty introduced by having to use values of model parameters that are not surely known. Finally **stochasticity** occurs when parameters or other quantities are not fixed but may vary.

A major difference in the way statisticians view modeling vs. that of risk modelers has to do with the complexity of the model. Very often models used in risk analysis are not simple empirical (i.e. data-based) models but mechanistic models. Mechanistic models are used due to the complexity of the problem. The choice of the model is a choice that leads to structural uncertainty. Discussion about structural uncertainty is often missing in statistical

texts or is connected to parameter uncertainty. For example, in multiple regression analysis structural uncertainty is associated with model misspecification. Model evaluation assumes a certain general structure (e.g. multiple linear) and the model is built through adding terms which are significant or which aid in prediction. Parameter uncertainty is typically discussed as a problem of estimation. Stochasticity is commonly dealt with as a measurement problem, under the assumption of an additive error. However, models incorporating stochasticity as a random effect are common, especially in the **analysis of variance**. These models are similar to models common to risk assessment, which often do not involve an error term. Stochasticity is often introduced through stochastic functions (e.g. weather) or random effects in parameter values.

Many modelers ignore the problem of structural uncertainty and focus on parameter and stochastic uncertainty. In theory, the problem of choice of model is satisfied by performing a 'validation' study, in which the goodness of fit of the model is assessed by comparing the results of the model with data. The importance of structural uncertainty analysis is in knowing when and where a model may be applied to produce reasonable results and, perhaps more importantly, where the model will fail.

It is more common in uncertainty analysis of complex models to focus on parameter and stochastic components of uncertainty. The reader of the risk analysis literature should be warned that there is inconsistent and varying terminology associated with these components. For example, uncertainty is sometimes associated only with parameter uncertainty, while stochasticity is referred to as variability, so authors discuss uncertainty and variability (e.g. [14]). Also, in an attempt to distinguish these components, the two components are sometimes called type A and B uncertainty. Type A uncertainty refers to variation in the objects that are measured or modeled. For example, if interest is in a model of the lead concentration in duck livers, then the basic unit is an individual duck. Variation may be caused by differential uptake rates (e.g. differences in uptake of lead in individual ducks). Type B uncertainty is associated with lack of knowledge of factors that are assumed constant across the objects (lack of knowledge of global parameters, etc.). In statistical models, type A uncertainty is analogous to within-treatment variability while type B is related to between-treatment variability. Stochasticity is generally viewed as uncertainty that is not reducible, while structural and parameter uncertainty are viewed as reducible (at least in principle) as more information is gathered.

Other views of uncertainty are quite possible and useful. For example, the fuzzy approach to uncertainty typically characterizes uncertainty in terms of measurement error, variability, and vagueness (*see* **Fuzzy numbers and nonprecise data**). The concept of vagueness leads to a nonprobabilistic view of uncertainty.

Statistics and Analysis of Uncertainty

To a large extent, the analysis of uncertainty is based on statistics and statistical thinking. In a loose sense, statistics is the science of collecting, analyzing, and interpreting information. The general model that forms the basis of much of classical statistical modeling is the model that posits that measurements may be interpreted as adding together a deterministic part with a random component:

$$\text{response} = \text{deterministic model} + \text{error} \qquad (1)$$

Thus, the response is modeled in terms of a deterministic model and a stochastic model or as an explained or certain component and an unexplained or uncertain component. In contrast to some risk assessments (e.g. population models), models used in statistical analysis are often simple. For example, the analysis of variance focuses on additive parameters (means or effects) and linear regression models involve terms which are additive and linear in the parameters ($\alpha + \beta x + \gamma x^2$ is valid, $\alpha + \alpha^2 x + \gamma x^2$ is not valid). The reason for the model simplicity is that *identifiability* is a common criterion for statistical models. A model is identifiable if its parameters can be estimated. Estimation is generally a simple task in the above models that are members of a broader class called **linear models**.

The statistical perspective is important, as it gives value to both the deterministic model and the error or stochastic component. The error term is representative of the uncertainty or variation in the response. The error term represents variation in the data not explained by the model and may contain several elements that are lumped together. The components of error include stochasticity, extraneous factors, or

terms not included in the model, either by choice or ignorance. Including error in the model allows us to model the error or uncertainty.

The models often considered are 'fixed effects' models. That is, the parameters are viewed as fixed but unknown values. Thus the model only contains deterministic elements and all stochastic or random elements are modeled as error. However, in the more general view, the model may contain both fixed and **random effects**. Thus both deterministic and stochastic components of the response may be modeled. For example, a general model from analysis of variance is

$$\text{response} = \text{deterministic component} + \text{stochastic component} + \text{error} \quad (2)$$

Here the stochastic part is usually modeled as a simple random variable, which adds only noise to the deterministic part (e.g. the random variable associated with the stochastic part has a normal distribution with mean 0 and unknown variance). The focus then is on estimating the unknown variance. These types of models are quite useful when unit-to-unit variation is of interest. The main application of these models has been where the unit is a human subject or an animal, although spatial units or laboratories could also be considered. Recent interest in statistics has focused on their use for applications involving non-normal data [6] (*see* **Generalized linear mixed models**).

The most common analysis of statistical uncertainty is based on the 'likelihood' function. The likelihood function is a function relating observations to parameters. Mathematically this approach is useful in estimation and testing problems. A common approach to estimation is to choose the values of parameters that make the data as likely as possible (maximize the function; *see* **Maximum likelihood estimation**). In testing, parameters may be evaluated by comparing the value of the likelihood with and without the parameter. The likelihood approach is the basis of most regression and analysis-of-variance analyses. For this case, the normal model is typically assumed. This model often leads to an analysis in terms of variance as the likelihood model reduces to a variance model.

The view just described is the classical or frequentist view of statistics. A different view is given in **Bayesian methods and modeling**. The difference between a Bayesian approach and a frequentist approach is the way that parameters in the model are treated. A frequentist treats the parameters as fixed quantities while the Bayesian allows the parameters to be random variables. For example, a common test in ecological risk analysis evaluates the mean chemical concentration in the field with a fixed 'safe' level. The frequentist views the mean concentration in the field as a fixed quantity; by taking a larger and larger sample, variation in this mean value decreases and the mean measured concentration becomes like a fixed value. The Bayesian approach says to treat the parameter of interest as a random value. Because the parameter is a random variable, it has a distribution. Thus, a Bayesian statistician may focus an analysis on the probability that the true mean concentration exceeds the safe concentration (*see* **Exceedance over threshold**). For a frequentist statistician, the true mean is either above or below the safe level and the probability is either one or zero.

Measuring Uncertainty

In statistical analysis, measurement of uncertainty is based on measures of the distribution that describes the uncertainty. The most common measure of uncertainty is *variance*. The variance of an estimated parameter describes how the parameter estimate would vary in repeated sampling. The variance of the estimate of risk estimates how much difference there would be in repeated estimates of risk. Other measures of uncertainty are based on distribution quantities such as upper and lower **quantiles**. When parameters are estimated, the likelihood may also be used to estimate uncertainty. The likelihood is associated with the data and how likely the observed data are, given the estimated parameter(s) (see [24]). The likelihood and measures based on the likelihood are commonly used to evaluate uncertainty in models.

Some Common Uncertainty Problems

There are many statistical problems in the analysis of uncertainty, especially for quantitative uncertainty analysis. Some general problems in uncertainty analysis which are briefly discussed below include:

- uncertainty factors;
- extrapolation models and prediction – estimation of the effects of uncertainty on model output and estimating uncertainty for unobserved systems (extrapolation);

- stochasticity – modeling the stochastic nature of the system and its effects;
- structural uncertainty and model specification – modeling relationships between variables with incomplete information;
- decision making with uncertain information;
- combining information as a method for reducing uncertainty;
- monitoring and methods for reducing uncertainty;
- estimation of uncertainty in complex models of risk (Monte Carlo methods);
- estimation of uncertainty using first-order analysis.

A more detailed list and discussion are given in [29], [32], or [19]. Quantitative approaches to the measurement of uncertainty vary with the complexity of the problem and the model that is used to evaluate risk. Some common types of uncertainty tools include uncertainty factors, extrapolation factors, statistical models, and Monte Carlo methods.

Uncertainty Factors. Risk analysis often relates exposure to a substance or agent with a measure of harm. A goal is often to find the level of the substance that results in no or minimal harm. An uncertainty factor – sometimes also called a safety factor – is a numerical quantity that is used to adjust the estimate of this level so that protection is ensured with high probability. The factor tries to adjust for uncertainties that occur when a risk analyst must make an adjustment for lack of information, often associated with attempting to extrapolate from one level of biological to a higher level of organization. Examples of such include [5, 32]:

- acute to chronic endpoints;
- response on individuals to response on populations;
- a single species response to a response in a group of species;
- laboratory to field conditions;
- one to many exposure routes;
- direct to indirect effects;
- a single location and time to multiple locations and/or times;
- a test organism to a human.

For example, data may be available on a single chemical and its toxicity to a single species from a short-term **biological assay**. The risk analyst may be more interested in long-term mortality. The effect of the chemical may be measured by the LC_{50}, the concentration of the chemical that results in 50% mortality. Given the lack of information to estimate the concentration that would result in similar mortality in the long-term test, the risk analyst uses the LC_{50} adjusted by a factor of 10. In a review of uncertainty factors, Duke and Taggart [9] found that although the approach was commonly used in risk assessment, the application of the factors was inconsistent.

Extrapolation Factors. Extrapolation factors refine the coarse quantitative model used for uncertainty factors by developing a model to extrapolate from one level to another. The uncertainty in the lower level is used to estimate the uncertainty in the level of interest. For example, suppose interest is in setting a safe level for a pesticide. The safe level might be defined as the level that protects the **ecosystem**. Information may not be available on the effects to the ecosystem but is available on laboratory studies of species that may or may not be present in the system. For example, there may be evidence from laboratory studies on toxicity to *Ceriodaphnia dubia* (a water flea), minnows, and brook trout. These taxa tend to survive in laboratory settings and thus are commonly used in toxicity testing (*see* **Aquatic toxicology**).

Each laboratory study produces an estimate of the 'endpoint' representing the effect of the toxicant, often expressed as a no-observable-effect level (*see* **Lowest-observed-adverse-effect level (LOAEL)**) or an LC_{50}. The goal of the assessment is to develop guidelines for the protection of the environment and is interpreted as the protection of all or most of the species in the system. This problem may now be viewed as a statistical problem. Since the endpoint is a measurement, it may be viewed as a random variable that has a distribution (typically a **lognormal** or log-logistic distribution) over the species in the ecosystem. Protection of the ecosystem may be viewed as protection of a certain percentage of the species in the ecosystem. Thus, if 95% of the species are protected (i.e. not severely affected by the pesticide at a particular concentration), then the ecosystem is said to be protected. The problem now becomes statistical: find the concentration of the pesticide that with high confidence will not

affect the chosen percentage of the species based on information in the sample of laboratory studies. This problem is essentially one of finding a tolerance limit. The estimate of a 'safe' level of the toxicant is given by

$$\log(\widehat{HC}_{\alpha,\gamma}) = \overline{X} - K_{\alpha,\gamma}s \tag{3}$$

where $HC_{\alpha,\gamma}$ is the hazardous concentration that protects $100(1-\alpha)\%$ of the species with confidence $100\gamma\%$, $\overline{X}$ is the mean of the endpoints from the laboratory studies, $K_{\alpha,\gamma}$ is the adjustment factor based on α, γ and the sample size, and s is the standard deviation of the measured endpoints.

A potential problem with the use of this interval is incorrect interpretation [28]. Confidence here means a degree of belief in the method. This means that if the HC values were computed, for example, 100 times (on 100 different random samples of species) for $\alpha = 0.05$ and $\gamma = 0.1$, then roughly 90% of those values would be protective of the true safety concentration (the concentration that protects 95% of the species). This method does not guarantee that the number computed from a single set of measurements is below the true safe concentration (i.e. does not guarantee safety). Furthermore, the method does not imply that the probability that the estimated safe concentration is below the true safe concentration is 0.90 (or, more generally, $100(1-\gamma)$).

The extrapolation method thus uses the variation in species sensitivity to calculate a concentration that is expected to be safe for most of the species present in the ecosystem. To be effective, the method relies on three assumptions:

1. Ecosystems are protected if $100(1-\alpha)\%$ of species are protected.
2. The distribution chosen to model variation in species sensitivities is correct.
3. The data are from randomly selected species and independent trials.

In reality, these assumptions are difficult to meet. For example, species in the ecosystem are rarely chosen for testing. This is because the species that will survive in laboratory conditions are limited and it is less expensive and simpler to use standard laboratory species. Applications of the method have not used a large number of species (five or more) so it is difficult to judge if the model of sensitivities is correct. Finally, ecosystems are complex entities with complex inter-relationships such as competition, **community food webs**, and predator–prey relationships. Combining information in a simple manner may not be appropriate for protection of the ecosystem (for additional concerns see [28]).

Uncertainty vs. Stochasticity. An important distinction that is often made in discussing and evaluating uncertainty is the difference between uncertainty and **stochasticity**. Some view stochasticity as a type of uncertainty while others view uncertainty separate from stochasticity. Others try to characterize uncertainty into type A and type B uncertainty. Part of the need to distinguish the different types of uncertainty arises from the desire to control uncertainty and categorize it. The ability to make quantitative distinction depends on the purpose of the assessment, the model used to evaluate risk, and the information that is available to evaluate uncertainty. For example, suppose that the analyst is interested in evaluating the risk associated with PCBs on bass. There may be several models considered to evaluate the risk. First, one may consider modeling the population of fish exposed to the PCBs. Second, one may consider a model of the effect on an individual fish. In both models, parameters need to be incorporated into the model. The quantities used in the model have uncertainty although they are viewed as fixed. This uncertainty if often viewed as uncertainty due to ignorance or lack of knowledge. Presumably, if enough information is collected, then these quantities may be calculated with minimal error and the uncertainty is reducible.

The second type of model is based on individuals in the population. In this case, there will still be parameters to evaluate. However, the parameters vary with the individual. In theory one might collect information on all individuals in the population; however, this is typically impossible and the more common view is that the individuals exhibit variation that may be modeled. The model would then involve a stochastic component to reflect how individual risk varies within the population.

Consider the example of a **water quality** standard for a chemical. The most basic view is that of a fixed standard or quantity which when exceeded signals a problem (*see* **Standards, environmental**). This view treats the standard as a fixed number and the results from a sample indicate a violation or no violation. This approach might be useful with chemicals such

as PCBs or pesticides sampled at a single site. Certain water quality measurement such as pH or dissolved oxygen show natural variability. Thus, in pristine sites, it may be possible for the measurement to exceed the fixed standard even though the site is not at risk. In this case, it may be valuable to consider the variation in the system and take a view that there is a distribution of values for the measurement. Then interest might focus on the probability of exceeding a fixed, biologically critical value.

In both of these cases, the value is a fixed value. If uncertainty in the value is also included then reducible uncertainty is involved. The uncertainty may be related to how risk to the environment is defined and measured.

Another example is where the measurement uncertainty is taken into account rather than the spatial variability (*see* **Spatial analysis in ecology**). The value of pH may have uncertainty due to location where it was sampled within a site as well as measurement error due to laboratory processing, handling, and equipment that is used. This later type would be viewed as reducible. Interpretation then involves determining if an uncertain value exceeds the standard. When both types of uncertainty are encountered, one needs to evaluate a probability of exceeding a standard when measurements are uncertain.

Decision Making Under Uncertainty. Like many statistical procedures, the assessment of risk may be viewed in terms of a decision procedure or an estimation procedure. In a decision procedure, there is often a sharp action such as whether or not a plant is impacting a site or if a chemical should be introduced. Decision processes are sometimes linked with hypotheses-testing procedures and the approach to testing is usually through a classical decision theory framework or a 'frequentist' viewpoint. In this view, the data are treated as the only information that is available for making the decision. Decisions are typically written as one of two hypotheses: a null hypothesis that represents 'no effect' and an alternative hypothesis which represents 'effect'.

Hypothesis testing is common in risk analysis. Examples include tests on whether or not a particular contaminant is hazardous, tests for the impact of a nuclear power plant (*see* **Nuclear risk**), and tests of models of effects (such as the relationship between dose and mortality). Several authors have indicated that there are several issues and uncertainties in the use of hypothesis tests [30]. Hypothesis testing requires assumptions about the data and model used to evaluate the relationship between stressor and effect. If uncertainties are not accounted for and reduced, then incorrect decisions are possible. Some problems with hypothesis testing include the influence of sample size and design on the power of the test. In addition, the basic testing model favors the null hypothesis, which is often not the hypothesis of interest [22], and the alternative hypothesis deserves more favor.

Several approaches are suggested for dealing with the problem. One approach is to use power analysis to be somewhat certain that the design and sample size are not overly influential. Another approach is to base the analysis on estimation rather than testing. One needs to remember, though, that estimation is also potentially problematic since sample size and design influence the variance of estimates.

Hypothesis testing and other methods need to be evaluated and developed for the problem they address. Part of a well-planned risk assessment involves data quality objectives (DQO). The US **Environmental Protection Agency** (EPA) has set a seven-step DQO process [31]. Issues that need to be addressed involve problem definition, spatial and temporal dimensions of the problem, the decision rule, and error rates that are acceptable. The DQO process must involve all stakeholders in the assessment to be effective. Part of the process is the setting of clear, defensible hypothesis tests. Hypothesis testing is not required in all risk assessments; in fact, estimation of risk is often more desirable. The hypothesis test forms one component of the assessment. The risk assessor should weigh all available information in the assessment and not rely on a single test.

A difficult problem in risk assessment is the interpretation of field data. Testing hypotheses with field data is problematic because the situation is not typical of designed experimental studies. For testing to be valid, it is required that the data are representative of the population of interest and measurements are independent.

Hypothesis testing in environmental regulatory situations has been criticized as favoring industry [22]. Null hypotheses are typically set as hypotheses of no effect. Lax quality standards, inadequate sample sizes, and high requirements for rejection of the null

hypothesis may lead to nonrejection of the hypothesis when in fact there is an effect. Careful consideration is required of both errors that are possible in decision analysis (false rejection and false acceptance). Situations in which one error rate is excessively large relative to the other should be avoided. Ideally, error rates that are acceptable should be set in the DQO process. If samples are to be collected to evaluate hypotheses, it is critical that the error rates be used to determine the sample size (i.e. a 'power' analysis [34]). Another possibility is to use bioequivalence tests [8] or to change the role of the null and alternative hypotheses.

In testing hypotheses an important consideration is the selection of the test. There are often several tests that may be applied to the same set of data. For example, in comparing a reference site with a suspected hazardous site, the two-sample *t*-test might be selected by some and the Wilcoxon test selected by others. Choice of statistic should depend on how the data relate to assumptions required for the validity of the test. The two-sample *t*-test is most efficient when the data come from a normal distribution. The Wilcoxon test does not require the data to be normally distributed (*see* **Ranks**). Both tests require independence of observations. Selecting the most appropriate statistical test is sometimes difficult and should be done in consultation with a statistician and other participants in the DQO process. This type of uncertainty is thus associated with the uncertainty due to the model that is chosen.

An alternative view is found in the Bayesian approach [3]. The process involves specification of the actions or decisions. These decisions are then related to parameters and a function describing the loss associated with the parameters is defined. A **prior distribution** is then specified for the parameters. This distribution describes information that we have about the parameter before we start the decision process. Given data, we can compute the distribution of the parameter adjusted for the information in the data (a **posterior distribution**) – using Bayes theorem. The posterior distribution represents the adjustment to and an update of the distribution of the prior by the data. Then the posterior distribution and the loss function are used to compute the expected loss associated with each decision. The loss function is typically chosen so the decision resulting in minimal loss is best.

The primary difference in the Bayesian view is that probabilities may be assigned to parameters. This allows for information to be incorporated through the prior distribution. This distribution is then altered by the data. The altered distribution is then used to make the decision. A difficult problem is the assignment of the prior distribution as it may vary among scientists. Hora and Iman [16] provide a formal approach for eliciting this information from experts (*see* **Elicitation**).

The Bayesian approach is also different in that it allows for multiple hypotheses. The method may be made more similar to a classical or frequentist test by computing **Bayes factors** [17]. Bayes factors are essentially measures of the strength of various hypotheses. The Bayes approach provides a more flexible method for assessment, especially when information about prior distributions is available and there are several competing hypotheses.

Structural Uncertainty. Structural uncertainty is perhaps the most overlooked aspect of uncertainty analysis. Most of the quantitative methodology for estimating uncertainty involves the assumption that the model chosen for risk estimation is correct. As will be pointed out by most ecologists, however, there are many instances in ecological risk assessment where we lack information about the model. This problem may be especially important in assessments, which evaluate risk over time or space as the model (or its components) may change. It is the responsibility of the modeler to explain the uncertainties associated with model specification and where possible to investigate the implications.

In the most general case, there will not be a solution to this problem, as uncertainty associated with surprises would have to be incorporated. A more tractable problem is associated with model variation. Certainly different modelers will prescribe different models. The variation in models can create a considerable dilemma for the risk manager and may also involve legal considerations. An interesting example is the case of striped bass in New York's Hudson River [1]. Some possible approaches are to consider, typically through simulation, the consequences of changes in the model [25]. Another useful approach is model mapping [26] which is useful for comparing inputs to different models. A Bayesian approach in which models are embedded in a larger class of models is also a possible approach [2]. A similar

Bayesian approach in simpler models is called model averaging [23]. In model averaging, results from different models are weighted and combined to produce an estimate averaged over the various models. Models are weighted based on their importance.

Combining Information. Risk analyses are rarely simple inferential processes. Information from related studies and related stressors may be valuable and should be included in the risk assessment. One approach for incorporation of other information is a **meta-analysis**. This approach is based on combining estimated quantities which have been obtained from similar, (usually) independent studies [13]. The approach can be quite powerful for combining information from the literature. However, even this approach is limited in that it tends to focus on single parameters from studies and this information is treated as the only information. Thus argument by analogy or the use of expert opinion is not allowed (see the discussion by Hanley [15] of the article by Gross [12], for an interesting example).

Another important area is in combining information from possibly unrelated studies. For example, the purpose of combining laboratory and field data studies is to gather information for unknown parameters and to make inferences on them using two sources of data. However, the problem is made difficult due to the need to combine different types of information in order to make a prediction. The linkages between two sources can be based on a correlation analysis for the variables between two sources, or certain prior distributions of the parameters from the two sources. Hierarchical Bayes models (*see* **Hierarchical Bayes methods**) can serve as a linkage model for this purpose [33].

Sampling Issues and Field Studies. Risk analysis is sometimes an iterative process. While the goal may be a decision, the process typically involves the collection and interpretation of data, possibly over a large period of time. For example, the risk assessment of the effects of power plants on striped bass in the Hudson River [1] involved 17 years of study and data collection. The collection and interpretation of field data for risk analysis is an important statistical problem and requires collaboration between statisticians and risk analysts for successful implementation. There are two additional problems involved.

First, there is the need for baseline monitoring data. This is required to increase understanding of basic ecological processes, for evaluating natural variation in systems with different degrees of disturbance, and for having prior data in case of anthropomorphic stress. It is not clear to what extent existing monitoring programs (EMEP, EMAP, etc.) will be useful for providing this information, but there is some concern that it is difficult to use sampling programs designed for one purpose in an alternative sampling program [20].

Second, there is the use of statistical sampling and design principles in the development of a monitoring program. Important problems include sample size for risk assessment studies, designing programs to detect changes of reasonable magnitude, and dealing with natural variation and trends as well as induced changes. Some other, less-discussed problems such as changes in funding and episodic effects are also important [29].

Monte Carlo Methods and the Estimation of Uncertainty. With most statistical methods, there is the assumption of a single distribution, which describes the uncertainty (i.e. the response is modeled using a Poisson, Gaussian, etc. distribution), or there is a common distribution, which is used to model uncertainty (i.e. the use of the normal model for random effects). This assumption greatly simplifies analysis as results may be expressed in terms of a single function, the likelihood function, or through parameters (mean and variance). While this approach is useful for some risk assessment work (especially laboratory analysis), general risk analysis is based on more complex models.

Uncertainty analysis of complex mathematical models of risk often involves the use of **simulation and Monte Carlo methods**. These methods are used to provide bounds or distributions on risk estimates. The method may assess parameter and/or stochastic uncertainty. The approach for assessing parameter uncertainty involves the following steps:

1. Select a distribution to describe possible values of a parameter.
2. Generate data from this distribution.
3. Use the generated data as possible values of the parameter in the model to produce output.

Although this is simple sounding, there are a number of potential problems that must be dealt with.

First, there is the selection (or elicitation) of the distribution. This may involve extensive work on the part of the risk analyst as the distribution describes the uncertainty about the parameter value. Often the distribution is based on the minimum, maximum, and mode of expected parameter values. It is impossible to exactly specify the distribution (e.g. lognormal vs. gamma). However, experience indicates that what is important is to choose distributions based on properties such as whether the distribution is skewed or symmetric, if it should be truncated or not, and whether extreme values should be allowed (see also [18]).

Some caution is of course warranted. In models involving time, there often is an assumption that the distribution does not change over time. Shlyakhter [27] has warned that we tend to be overly optimistic about the selected distributions, as unsuspected errors are not accounted for and that uncertainty is often underestimated. Other potential problems include correlations between parameters, correlation structures over time [10], model parameterization problems (different parameterizations of the same model may lead to different results with respect to uncertainty), and lack of identifiability. Further work is needed in this area before sharp guidelines can be developed. Burmaster and Anderson [4] provide some guidelines for proper assessment of uncertainty using Monte Carlo methods.

Another complexity is the model of uncertainty. In the above scenario, the only uncertainty arises from the choice of parameters. In applications, we might consider uncertainty due to the units of the model. This might be humans in a human health risk assessment or **lakes** in an assessment of **acid rain** risk. Uncertainty associated with a quantity that is fixed is referred to (above) as type B uncertainty (parameters). When the quantity is random there is a second uncertainty that results which is type A uncertainty (humans, lakes). In this case, the assessment endpoint is often a distribution of values rather than a single value. Thus, type A uncertainty would lead to a distribution of the endpoint of interest while type B uncertainty is associated with how the distribution changes. An example would be a situation in which the endpoint varies over space. The endpoint would have a distribution over an area but the distribution would have uncertainty due, say, to unknown parameters. Thus, the uncertainty model would have to include both a model for uncertainty in parameters as well as units.

An Illustrative Example

It is illustrative to compare Monte Carlo uncertainty analysis with a standard statistical assessment to better understand the method and different types of uncertainties. As a simple example, consider the simple linear regression model with no error. This might represent the risk model associated with chlorophyll a as a function of phosphorus. Thus we have that

$$\text{chlorophyll } a = \beta_0 + \beta_1 \text{phosphorus} \tag{4}$$

or more generally

$$y = \beta_0 + \beta_1 x \tag{5}$$

This model is a deterministic model; the intercept and slope coefficients are viewed as fixed. The true parameters are unknown and would have to be estimated or assigned values. Thus for prediction of chlorophyll a, the model that is used can be described by

$$\hat{y} = b_0 + b_1 x \tag{6}$$

For an uncertainty analysis associated with the parameters, we would assume that the parameters have distributions. Consider three possible cases. The first is a random slope model in which b_0 is known and b_1 has an $N(\beta_1, \sigma_1^2)$ distribution. The second is a model with both the slope and intercept having a random distribution, in which b_0 is $N(\beta_0, \sigma_0^2)$, and b_1 is $N(\beta_1, \sigma_1^2)$. Finally the parameters are allowed to have a joint distribution: b_0, b_1 are bivariate $N[(\beta_0, \beta_1)', V]$, where V is the variance–covariance matrix (*see* **Bivariate distributions**).

In an application, we would have to have values of the parameter. These may be obtained from data collected at a particular site or from former studies. To relate the analysis to standard regression analysis, suppose that we use the data in Table 1, collected on the two variables of interest, and fit the statistical model

$$y = \beta_0 + \beta_1 x + \varepsilon \tag{7}$$

where ε is $N(0, \sigma^2)$.

Fitting the data in Table 1, the estimated model is $\hat{y} = -1.47 + 1.59x$, with variance estimated as 1.21.

Table 1 Twenty-two observations used to model the relationship between chlorophyll *a* and total phosphorus. Values are log transformed

Observation	Log total phosphorus	Log chlorophyll *a*
1.00	1.97	1.92
2.00	2.07	2.36
3.00	2.45	2.64
4.00	2.55	1.17
5.00	2.77	2.07
6.00	2.93	2.22
7.00	3.30	3.78
8.00	3.60	6.30
9.00	3.65	4.59
10.00	3.96	3.02
11.00	3.79	6.30
12.00	4.23	5.64
13.00	4.43	5.78
14.00	4.65	7.00
15.00	4.91	4.67
16.00	4.94	7.40
17.00	5.18	6.80
18.00	5.52	5.75
19.00	5.59	8.37
20.00	6.01	7.90
21.00	5.90	7.93

Values of the slope and intercept are now drawn from normal distributions. The variance in the normal distribution is chosen to coincide with that from the regression model. For the intercept, the distribution is N(−1.47, 0.67) while for the slope the distribution N(1.59, 0.038) was used. The values for the variances correspond to the values from the regression output. For the last scenario, the covariance matrix was used. This matrix included the variance for the parameters and the covariance. The covariance used a correlation of 0.95.

Figures 1–3 display 100 lines from the model with the three different scenarios. Note that when the correlation is not accounted for (slope and intercept are treated as independent), the lines seem to vary widely. When the correlation is added the variation is greatly reduced and we obtain a picture similar to what is expected in a regression analysis. The figures look like what is obtained through prediction bands on the regression line. To compare the figures with regression bands, we can compute these from the regression results.

Suppose that the interest is in obtaining information on the mean value of log(chlorophyll *a*) at log(phosphorus) or $x = 4$. The 95% confidence interval for μ_y at $x = 4$ is (4.40, 5.41) with prediction variance 0.241. The 95% prediction interval for a new observation is (2.54, 7.26).

A similar interval can be obtained from Monte Carlo methods by calculating log(chlorophyll *a*) at log(phosphorus) = 4 from all of the simulated regressions. This was done using the program Crystal Ball [7] with 10 000 simulations and the distributions are displayed in Figures 4–6. A 95% confidence interval is obtained by finding the 2.5th and 97.5th percentiles. For the last model (Figure 6) the resulting 95% interval is (4.39, 5.40). This is essentially the same as what was obtained using the statistical theory. Thus what was done in the third Monte Carlo run was to use simulation to calculate a confidence interval for the mean value of log(chlorophyll *a*). Note that the other Monte Carlo runs do not produce the same interval and in fact produce intervals that

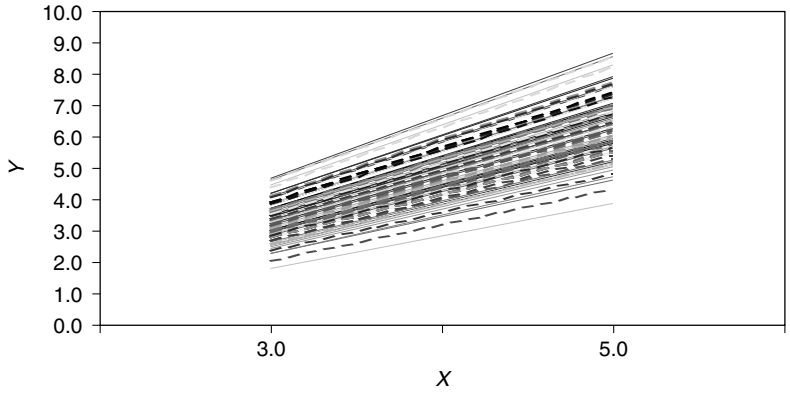

Figure 1 Sample lines assuming uncertainty only in the slope

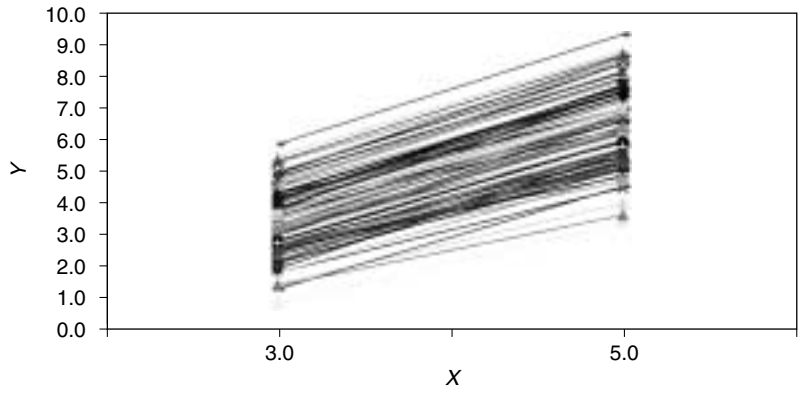

Figure 2 Sample lines assuming uncertainty in slope and intercept

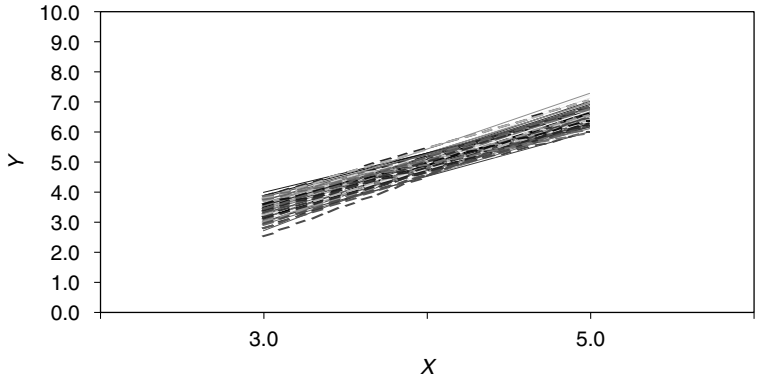

Figure 3 Sample lines assuming uncertainty in slope and intercept with correlation

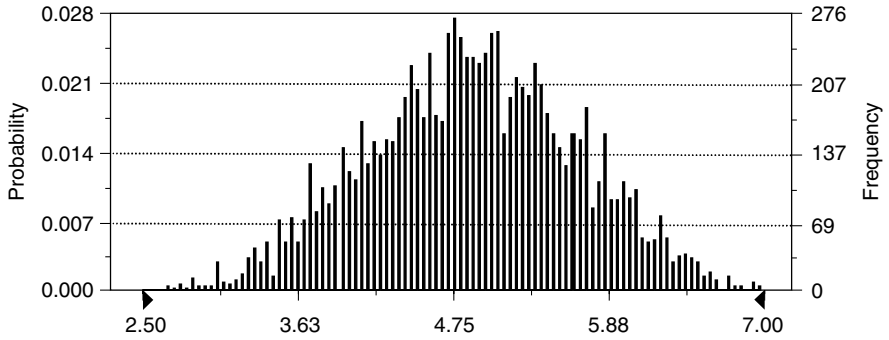

Figure 4 Predictions for the simulation in which the slope varies

are considerably wider. Also, the interval represents a confidence interval rather than a prediction interval. To obtain a prediction interval, one would add an error term to the model with variance from the regression model.

This example illustrates several points that are important to consider in planning, interpreting, and evaluating a Monte Carlo analysis. First, choosing what parameters to model is important. Clearly the results differ considerably when different parameters

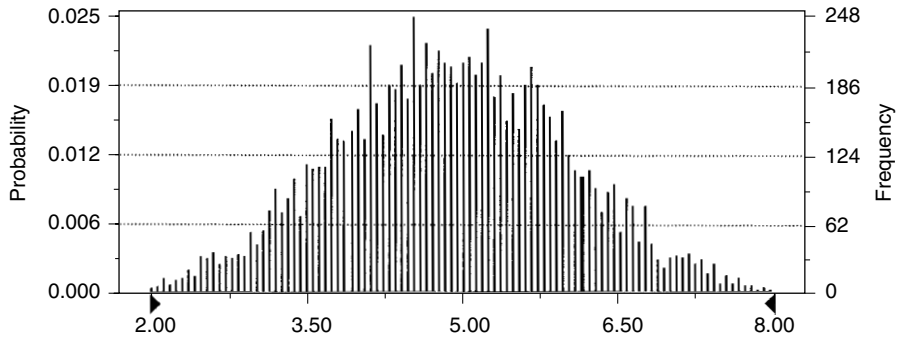

Figure 5 Predictions for the simulation in which the slope and intercept vary but are uncorrelated

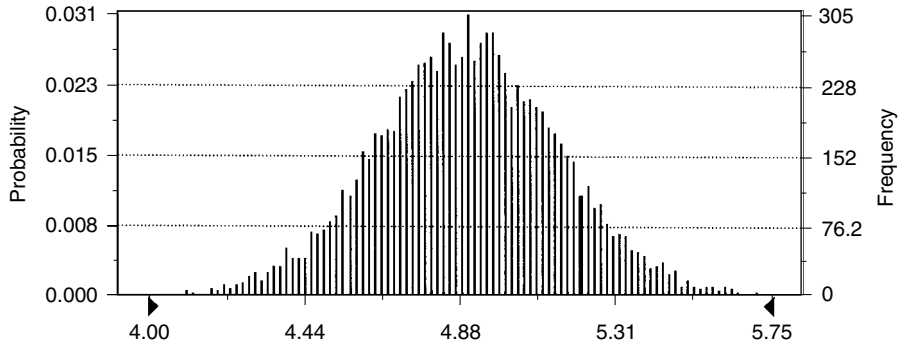

Figure 6 Predictions for the simulation in which the slope and intercept vary and are correlated

are varied. Second, the distribution chosen to model the uncertainty is important. It is important to recognize the difference between distributions associated with parameter uncertainty and stochasticity. Third, the correlation structure may be quite important. Ignoring the correlation may lead to inappropriate intervals and distributions. Interpretation of the output depends on the input. Interpretation is different when stochasticity is part of the analysis. Decisions need to be made early in the risk assessment about the model and whether the focus is on individuals or populations. A final point to consider is that the Monte Carlo method is fairly complex even for a model such as multiple regression. More complicated models are common in risk assessment and require careful planning.

Latin Hypercube Sampling

Models that contain a large number of parameters can be difficult to analyze using Monte Carlo methods. The difficulty arises in trying to obtain a representative sample of parameter values from their distributions. Having a large number of parameters requires a large number of Monte Carlo simulations to produce defensible results. One approach to obtaining a representative set of model output is to use Latin hypercube sampling to obtain a representative set of input samples to evaluate using the model.

Latin hypercube sampling is based on stratified sampling of probability distributions. The approach is to divide each distribution into n equal probability intervals where n is the number of simulations. For each run of the simulation one parameter value is selected from each distribution, each region being selected only once. Thus if each distribution is divided into two parts and there are p parameters, there will be 2^p possible sets of regions to select from. If a random approach is used for selection, then a point is selected from each interval. An alternative is to use the median of each interval in the analysis.

Latin hypercube sampling generally gives better results in calculating the tails of the distribution of risk and requires fewer simulations relative to

ordinary Monte Carlo analysis. This may be of considerable importance when the number of parameters is large.

Sensitivity Analysis

A tool related to uncertainty analysis is **sensitivity analysis**. Sensitivity analysis is used to determine the importance of different parameters and components of the model on the output of the model. If the response variable y depends on several variables, then the sensitivity of the response with respect to the variable or parameter is measured by the derivative of the response with respect to the variable or parameter.

Sensitivity analysis is sometimes a by-product of a Monte Carlo uncertainty analysis. For example, if interest is in the sensitivity of the response to changes in variables, the values of the variables are selected using a probability method and then run through the model. The result is a set of input and output quantities. The importance of a variable is measured by the correlation or partial correlation between the variable and the response. Variables with the greatest (positive or negative) correlation indicate variables with great sensitivity.

First-order Analysis

In first-order analysis a model is linearized and then the uncertainty is measured in terms of the variance. The variable y is considered an output variable that is related to the input variables, the xs, through the model. The xs are measured quantities and information is known about their variances. Then, the variance associated with y is approximately given by

$$\sigma_y^2 \approx \left(\frac{\partial y}{\partial x_1}\right)^2 \sigma_{x_1}^2 + \left(\frac{\partial y}{\partial x_2}\right)^2 \sigma_{x_2}^2 + \cdots + \left(\frac{\partial y}{\partial x_p}\right)^2 \sigma_{x_p}^2 \tag{8}$$

when the variables are independent. When the variables are dependent, covariances must be included.

In applications, one may calculate the quantities directly when the models are simple or use regression methods with Monte Carlo analysis to compute the quantities involved. Variables or parameters are varied using Monte Carlo, resulting in an array of y and x values. The values are then analyzed using regression analysis. The analysis may use multiple regression analysis with only simple terms or more complex regression would use a response surface approach. The multiple regression approach would ignore interactions (correlations) between the variables. The slopes from the multiple regression model would be used to estimate the derivatives.

Other Approaches for Assessing Uncertainty

There are a number of other approaches for evaluating uncertainty including Bayesian uncertainty analysis [21], fuzzy logic, interval analysis, and Laplace and Mellin transformations. Fuzzy logic is a method that substitutes degrees of belief for probability and tries to mathematically quantify vague concepts such as 'pretty warm'. Interval analysis is a mathematical tool that deals with intervals and the mathematics of intervals. Further discussion of these methods may be found in [10] or [32].

References

[1] Barnthouse, L.W., Klauda, R.J., Vaughn, D.S. & Kendall, R.L., eds (1988). *Science, Law, and Hudson River Power Plants: A Case Study in Environmental Impact Assessment, American Fisheries Society Monograph*, Vol. 4, American Fisheries Society, Bethesda.

[2] Bernier, J.B. (1991). Bayesian analysis of robustness of models in water and environmental sciences, in *Water Resources in Engineering Risk Assessment*, P. Ganoulis, ed., Springer-Verlag, Berlin.

[3] Box, G.E.P. & Tiao, G.C. (1973). *Bayesian Inference in Statistical Analysis*, Wiley, New York.

[4] Burmaster, D.E. & Anderson, P.D. (1994). Principles of good practice for the use of Monte Carlo techniques in human health and ecological risk assessments, *Risk Analysis* **14**, 477–481.

[5] Calabrese, E.J. & Baldwin, L.A. (1993). *Performing Ecological Risk Assessments*, Lewis, Chelsea.

[6] Davidian, M. & Giltinan, D.M. (1995). *Nonlinear Models for Repeated Measurement Data*, Chapman & Hall, London.

[7] Decisioneering Inc. (2000). *Crystal Ball 2000 Users Manual*, Decisioneering, Denver.

[8] Dixon, P.M. & Garrett, K.A. (1994). Statistical issues for field experimenters, in *Wildlife Toxicology and Population Modeling: Integrated Studies of Agroecosystems*, Lewis, Boca Raton, pp. 439–450.

[9] Duke, L.D. & Taggart, M. (2000). Uncertainty factors in screening ecological risk assessments, *Environmental Toxicology and Chemistry* **19**, 1668–1680.

[10] Ferson, S. & Long, F. (1994). Conservative uncertainty propagation in environmental risk assessments, in *Environmental Toxicology and Risk Assessment, ASTM STP 1218*, Vol. 3, J.S. Hughes, G.R. Biddinger & E. Mones, eds, American Society for Testing and Materials, Philadelphia.

[11] Good, I.J. (1994). Reliability always depends on probability of course, *Journal of Statistical Computation and Simulation* **52**, 192–193.

[12] Gross, A.J. (1995). Uncertainties in lung cancer risk estimates reported for exposure to environmental tobacco smoke, *Environmetrics* **6**, 401–410.

[13] Gurevitch, J. & Hedges, L.V. (1993). Meta-analysis: combining the results of independent experiments, in *Design and Analysis of Ecological Experiments*, S.M. Scheiner & J. Gerevitch, eds, Chapman & Hall, London, pp. 378–398.

[14] Haimes, Y.Y., Barry, T. & Lambert, J.H. (1994). When and how can you specify a probability distribution when you don't know much? Workshop Proceedings, *Risk Analysis* **14**, 661–704.

[15] Hanley, J.A. (1995). Uncertainties in lung cancer risk estimates reported for exposure to environmental tobacco smoke – a rejoinder to Gross, *Environmetrics* **6**, 419–422.

[16] Hora, S.C. & Iman, R.L. (1989). Expert opinion in risk analysis: the NUREG-1150 methodology, *Nuclear Science and Engineering* **102**, 323–331.

[17] Kass, R.E. & Raftery, A.E. (1995). Bayes factors, *Journal of the American Statistical Association* **90**, 773–795.

[18] Lambert, J.H., Matalas, N.C., Ling, C.W., Haimes, Y.Y. & Li, D. (1994). Selection of probability distributions in characterizing risk of extreme events, *Risk Analysis* **14**, 731–743.

[19] Morgan, M.G. & Henrion, M. (1990). *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*, Cambridge University Press, Cambridge.

[20] Overton, J., Young, T.C. & Overton, W.S. (1993). Using 'found' data to augment a probability sample: procedure and case study, *Environmental Monitoring and Assessment* **26**, 65–83.

[21] Patwardhan, A. & Small, M.J. (1992). Bayesian methods for model uncertainty analysis with applications to future sea level rise, *Risk Analysis* **12**, 513–523.

[22] Peterman, R. (1990). Statistical power analysis can improve fisheries research and management, *Canadian Journal of Fisheries and Aquatic Sciences* **47**, 2–15.

[23] Raftery, A.E. (1995). Bayesian model selection in social research (with discussion by Andrew Gelman, Donald B. Rubin & Robert M. Hauser), in *Sociological Methodology*, P.V. Marsden, ed., Blackwell, Oxford, pp. 111–196.

[24] Ramsey, F.L. & Schafer, D.W. (1997). *The Statistical Sleuth: A Course in the Methods of Statistics*, Duxbury Press, Belmont.

[25] Rose, K.R., Brenkert, A.L., Cook, R.B. & Gardner, R.H. (1991). Systematic comparison of ILWAS, MAGIC and ETD watershed acidification models 2: Monte Carlo under regional variability, *Water Resources Research* **27**, 2591–2603.

[26] Rose, K.R., Cook, R.B., Brenkert, A.L. & Gardner, R.H. (1991). Systematic comparison of ILWAS, MAGIC and ETD watershed acidification models 1: Mapping among model inputs and deterministic results, *Water Resources Research* **27**, 2577–2589.

[27] Shlyakhter, A. (1994). An improved framework for uncertainty analysis: accounting for unsuspected errors, *Risk Analysis* **14**, 441–447.

[28] Smith, E.P. & Cairns, J.J. (1993). Extrapolation methods for setting ecological standards for water quality: statistical and ecological concerns, *Ecotoxicology* **2**, 203–219.

[29] Smith, E.P. & Shugart, H.H. (1994). *Uncertainty in Ecological Risk Assessment*, US EPA, Washington, pp. 8-1–8-53.

[30] Suter, G.W.I. (1993). *Ecological Risk Assessment*, Lewis, Boca Raton.

[31] US Environmental Protection Agency (1994). *Guidance for Data Quality Objectives Process*, US EPA, Quality Assurance Management Staff, Washington.

[32] Warren-Hicks, W.J. & Moore, D.R.J., eds (1998). *Uncertainty Analysis in Ecological Risk Assessment, Proceedings from the Pellston Workshop on Uncertainty Analysis in Ecological Risk Assessment*, Society of Environmental Chemistry and Toxicology (SETAC), Pensacola, Pellston, August 1995, pp. 23–28.

[33] Wolpert, R.L., Reckhow, K.H. & Steinberg, L.J. (1992). Bayesian hierarchical models for transport and fate modeling, in *Bayesian Statistics in Science and Technology: Case Studies*, J.S. Hodges et al., eds, Springer-Verlag, New York, pp. 241–296.

[34] Zar, J. (1984). *Biostatistical Analysis*, 2nd Edition, Prentice-Hall, Englewood Cliffs.

(*See also* **Model uncertainty**; **Risk assessment, probabilistic**; **Risk perception**)

ERIC SMITH