9. Model Sensitivity and Uncertainty Analysis

9 Model Sensitivity and Uncertainty Analysis

The usefulness of any model depends in part on the accuracy and reliability of its output. Yet, because all models are imperfect abstractions of reality, and because precise input data are rarely if ever available, all output values are subject to imprecision. Input data errors and modelling uncertainties are not independent of each other – they can interact in various ways. The end result is imprecision and uncertainty associated with model output. This chapter focuses on ways of identifying, quantifying, and communicating the uncertainties in model outputs.

1. Introduction

Models are the primary way we have to estimate the multiple effects of alternative water resources system design and operating policies. Models predict the values of various system performance indicators. Their outputs are based on model structure, hydrological and other time-series inputs, and a host of parameters whose values describe the system being simulated. Even if these assumptions and input data reflect, or are at least representative of, conditions believed to be true, we know they will be inaccurate. Our models are always simplifications of the real systems we study. Furthermore, we simply cannot forecast the future with precision, so we know the model outputs of future conditions are at best uncertain.

Some prediction uncertainties can be reduced by additional research and data collection and analysis. Before undertaking expensive studies to gather and analyse additional data, it is reasonable to ask what improvement in estimates of system performance, or what reduction in the uncertainty associated with those estimates, would result if all data and model uncertainties could be reduced. Such information helps determine how much one would be willing to 'pay' in order to reduce prediction uncertainty. If prediction uncertainty is costly, it may pay to invest in additional data collection, more studies or better models – all

aimed at reducing that prediction uncertainty. If that uncertainty has no, or only a very modest, impact on the likely decision that is to be made, one should find other issues to worry about.

If it appears that reducing prediction uncertainty is worthwhile, then one should consider how best to do it. If it involves obtaining additional information, then it is clear that the value of this additional information, however measured, should exceed the cost of obtaining it. The value of such information will be the increase in system performance, or the reduction in its variance, that one can expect from obtaining such information. If additional information is to be obtained, it should be information that reduces the uncertainties considered important, not the unimportant ones.

This chapter reviews some methods for identifying and communicating model prediction uncertainty. The discussion begins with a review of the causes of risk and uncertainty in model output. It then examines ways of measuring or quantifying uncertainty and model output sensitivity to model input imprecision, concentrating on methods that seem most relevant or practical for large-scale regional simulation modelling. It builds on some of the statistical methods reviewed in Chapter 7 and the modelling of risk and uncertainty in Chapter 8.

2. Issues, Concerns and Terminology

Outcomes or events that cannot be predicted with certainty are often called risky or uncertain. Some individuals draw a special and interesting distinction between risk and uncertainty. In particular, the term *risk* is often reserved to describe situations for which probabilities are available to describe the likelihood of various events or outcomes. If probabilities of various events or outcomes cannot be quantified, or if the events themselves are unpredictable, some would say the problem is then one of *uncertainty*, and not of risk. In this chapter, what is not certain is considered uncertain, and uncertainty is often described by a probability distribution. When the ranges of possible events are known and their probabilities are measurable, risk is called *objective risk*. If the probabilities are based solely on human judgement, the risk is called *subjective risk*.

Such distinctions between objective and subjective risk, and between risk and uncertainty, rarely serve any useful purpose to those developing and using models. Likewise, the distinctions are often unimportant to those who should be aware of the risks or uncertainties associated with system performance indicator values.

Uncertainty in information is inherent in future-oriented planning efforts. It stems from inadequate information and incorrect assumptions, as well as from the variability of natural processes. Water managers often need to identify both the uncertainty and the sensitivity of, or changes in, system performance indicator values due to any changes in possible input data and parameter values from what were predicted. They need to reduce this level of uncertainty to the extent practicable. Finally, they need to communicate the residual uncertainties clearly so that decisions can be made with this knowledge and understanding.

Sensitivity analysis can be distinguished from uncertainty analysis. Sensitivity analysis procedures explore and quantify the impact of possible errors in input data on predicted model outputs and system performance indices. Simple sensitivity analysis procedures can be used to illustrate either graphically or numerically the consequences of alternative assumptions about the future. Uncertainty analyses employing probabilistic descriptions of model inputs can be used to derive probability distributions of model outputs and system performance indices. Figure 9.1 illustrates the impact of both input data sensitivity and input data uncertainty on model output uncertainty.

It is worthwhile to explore the transformation of uncertainties in model inputs and parameters into uncertainty in model outputs when conditions differ from those reflected by the model inputs. Historical records of system characteristics are typically used as a basis for model inputs. Yet conditions in the future may change. There

Figure 9.1. Schematic diagram showing relationship among model input parameter uncertainty and sensitivity to model output variable uncertainty (Lal, 1995).
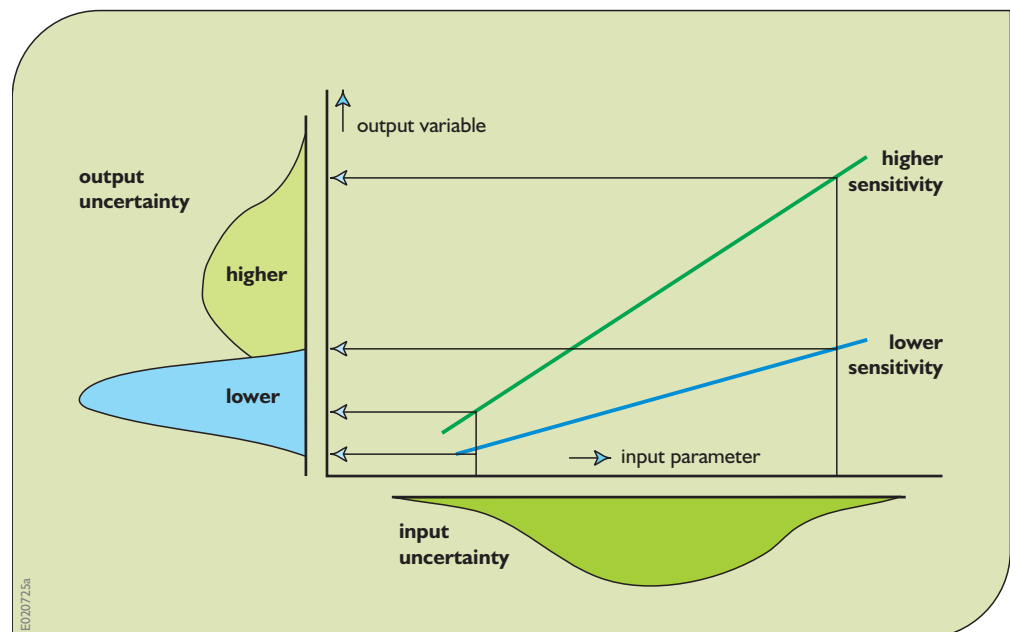
may be changes in the frequency and amounts of precipitation, changes in land cover and topography, and changes in the design and operation of control structures, all resulting in changes of water stages and flows, and their qualities, and consequently changes in the affected ecosystems.

If asked how the system would operate with inputs similar to those in the historical database, the model should be able to interpolate within the available knowledge base to provide a fairly precise estimate. Still, that estimate will not be perfect. This is because our ability to reproduce current and recent operations is not perfect, though it should be fairly good. If asked to predict system performance for situations very different from those in the historical knowledge base, or when the historical data are not considered representative of what might happen in the future (due to climate change for example), such predictions become much less precise.

There are two reasons for this. First, our description of the characteristics of those different situations or conditions may be imprecise. Second, our knowledge base may not be sufficient to calibrate model parameters in ways that would enable us to reliably predict how the system will operate under conditions unlike those that have been experienced historically. The more conditions of interest differ from those in the historical knowledge base, the less confidence we have that the model is providing a reliable description of systems operation. Figure 9.2 illustrates this issue.
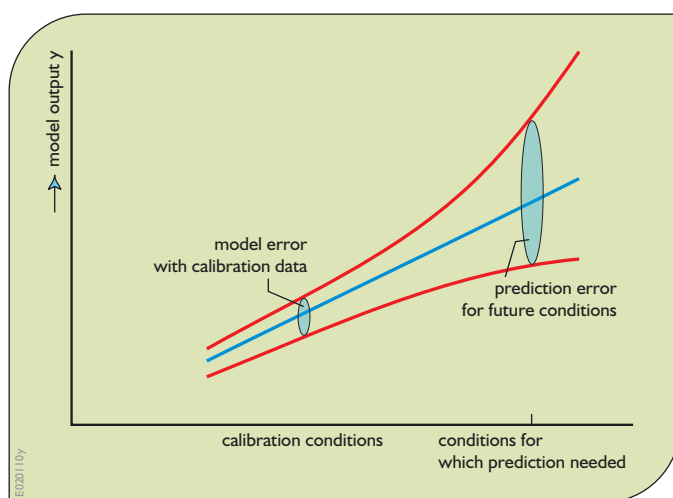
Figure 9.2. The precision of model predictions is affected by the difference between the conditions or scenarios of interest and the conditions or scenarios for which the model was calibrated.

Clearly, an uncertainty analysis needs to consider how well a model can replicate current operations, and how similar the target conditions or scenarios are to those described in the historical record. The greater the required extrapolation from what has been observed, the greater will be the importance of parameter and model uncertainties.

The relative and absolute importance of different parameters will depend on the system performance indicators of interest. Seepage rates may have a very large local impact, but a small global effect. Changes in system-wide evapotranspiration rates are likely to affect system-wide flows. The precision of model projections and the relative importance of errors in different parameters will depend upon:

- the precision with which the model can reproduce observed conditions
- the difference between the predicted conditions and the historical experience included in the knowledge base
- system performance characteristics of interest.

Errors and approximations in input data measurement, parameter values, model structure and model solution algorithms are all sources of uncertainty. While there are reasonable ways of quantifying and reducing these errors and the resulting range of uncertainty of various system performance indicator values, they are impossible to eliminate. Decisions will still have to be made in the face of a risky and uncertain future, and can be modified as new data and knowledge are obtained in a process of adaptive management.

There is also uncertainty with respect to human behaviour and reactions related to particular outcomes and their likelihoods, that is, to their risks and uncertainties. As important as risks and uncertainties associated with human reactions are to particular outcomes, they are not usually part of the models themselves. Social uncertainty may often be the most significant component of the total uncertainty associated with just how a water resources system will perform. For this reason we should seek designs and operating policies that are flexible and adaptable.

When uncertainties associated with system operation under a new operating regime are large, one should anticipate the need to make changes and improvements as experience is gained and new information accumulates. When predictions are highly unreliable, responsible

managers should favour actions that are robust (good under a wide range of situations), gain information through research and experimentation, monitor results to provide feedback for the next decision, update assessments and modify policies in the light of new information, and avoid irreversible actions and commitments.

3. Variability and Uncertainty in Model Output

Differences between model output and observed values can result from either natural variability, such as is caused by unpredictable rainfall, evapotranspiration, water consumption and the like, and/or by both known and unknown errors in the input data, the model parameters or the model itself. The later is sometimes called knowledge uncertainty, but it is not always due to a lack of knowledge. Models are always simplifications of reality and, hence, 'imprecision' can result. Sometimes imprecision occurs because of a lack of knowledge, such as just how a particular species will react to various environmental and other habitat conditions. At other times, known errors are introduced simply for practical reasons.

Imperfect representation of processes in a model constitutes model *structural uncertainty*. Imperfect knowledge of the values of parameters associated with these processes constitutes model *parameter uncertainty*. *Natural variability* includes both *temporal variability* and *spatial variability*, to which model input values may be subject.

Figure 9.3 illustrates these different types of uncertainty. For example, the rainfall measured at a weather station within a particular model grid cell may be used as an input value for that cell, but the rainfall may actually vary at different points within that cell and its mean value will vary across the landscape. Knowledge uncertainty can be reduced through further measurement and/or research. Natural variability is a property of the natural system, and is usually not reducible at the scale being used. Decision uncertainty is simply an acknowledgement that we cannot predict ahead of time just what decisions individuals and organizations will make, or even just what particular set of goals or objectives will be considered and the relative importance of each.

Rather than contrasting 'knowledge' uncertainty versus natural variability versus decision uncertainty, one can classify uncertainty in another way based on specific sources of uncertainty, such as those listed below, and address ways of identifying and dealing with each source of uncertainty.

- Informational uncertainties
  - imprecision in specifying the boundary and initial conditions that impact the output variable values
  - imprecision in measuring observed output variable values.
- Model uncertainties
  - uncertain model structure and parameter values
  - variability of observed input and output values over a region smaller than the spatial scale of the model
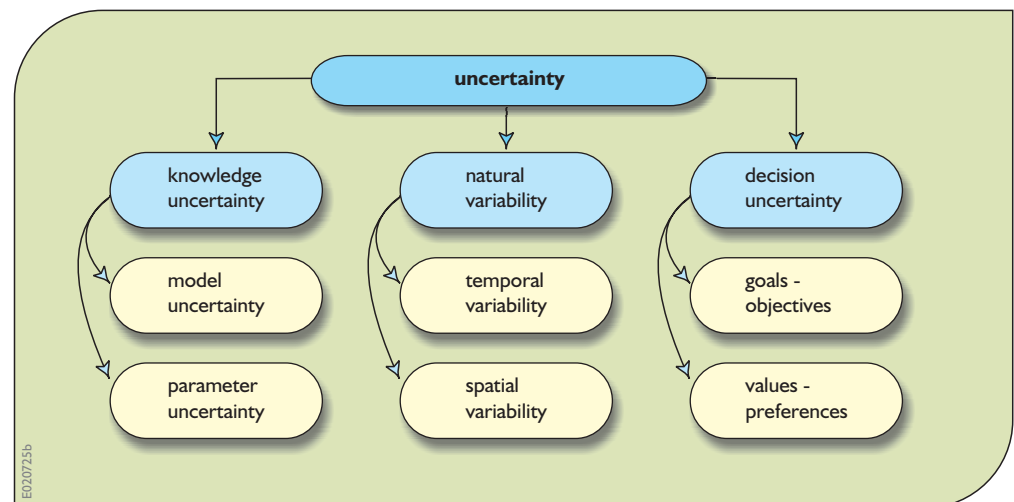
Figure 9.3. One way of classifying types of uncertainty.

– variability of observed model input and output values within a time smaller than the temporal scale of the model. (e.g., rainfall and depths and flows within a day)
– errors in linking models of different spatial and temporal scales.
• Numerical errors
– errors in the model solution algorithm.

3.1. Natural Variability

The main source of hydrological model output value variability is the natural variability in hydrological and meteorological input series. Periods of normal precipitation and temperature can be interrupted by periods of extended drought and intense meteorological events such as hurricanes and tornadoes. There is no reason to believe that such events will not continue to occur and become even more frequent and extreme. Research has demonstrated that climate has been variable in the past and there are concerns about anthropogenic activities that may increase that variability each year. Sensitivity analysis can help assess the effect of errors in predictions if those predictions are based only on past records of historical time-series data describing precipitation, temperature and other exogenous forces across and on the border of the regions being studied.

Time-series input data are often actual, or at least based on, historical data. The time-series values typically describe historical conditions including droughts and wet periods. What is distinctive about natural uncertainty, as opposed to errors and uncertainty due to modelling limitations, is that natural variability in meteorological forces cannot be reduced by improving the model's structure, increasing the resolution of the simulation or better calibration of model parameters.

Errors result if meteorological values are not measured or recorded accurately, or if mistakes are made in the generation of computer data files. Furthermore, there is no assurance that the statistical properties of historical data will accurately represent those of future data. Actual future precipitation and temperature scenarios will be different from those in the past, and this difference in many cases may have a larger effect than the uncertainty due to incorrect parameter values. However, the effects of uncertainties in the parameter values used in stochastic
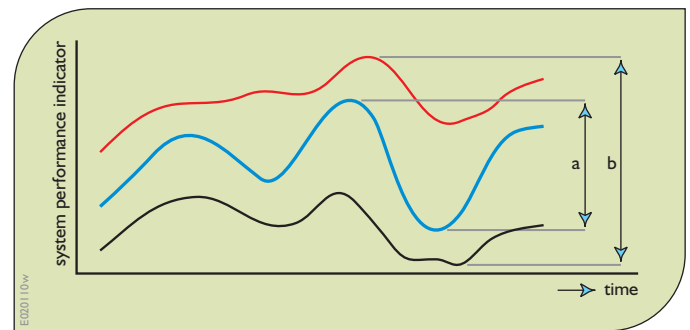

Figure 9.4. Time series of model output or system performance showing variability over time. Range 'a' results from the natural variability of input data over time. The extended range 'b' results from the variability of natural input data as well as from imprecision in input data measurement, parameter value estimation, model structure and errors in model solution algorithms. The extent of this range will depend on the confidence level associated with that range.

generation models are often much more significant than the effects of using different stochastic generation models (Stedinger and Taylor, 1982).

While variability of model output is a direct result of variability of model input (e.g. hydrological and meteorological data), the extent of the variability, and its lower and upper limits, may also be affected by errors in the inputs, the values of parameters, initial boundary conditions, model structure, processes and solution algorithms.

Figure 9.4 illustrates the distinction between the variability of a system performance indicator due to input data variability, and the extended range of variability due to the total uncertainty associated with any combination of the causes listed in the previous section. This extended range is what is of interest to water resources planners and managers.

What can occur in practice is a time series of system performance indicator values that can range anywhere within or even outside the extended range, assuming the confidence level of that extended range is less than 100%. The confidence one can have that some future value of a time series will be within a given range is dependent on two factors. The first is the number of measurements used to compute the confidence limits. The second is the assumption that those measurements are representative (come from the same statistical or stochastic process yielding) future measurements, i.e. they come from the same statistical or stochastic process. Figure 9.5 illustrates this point.
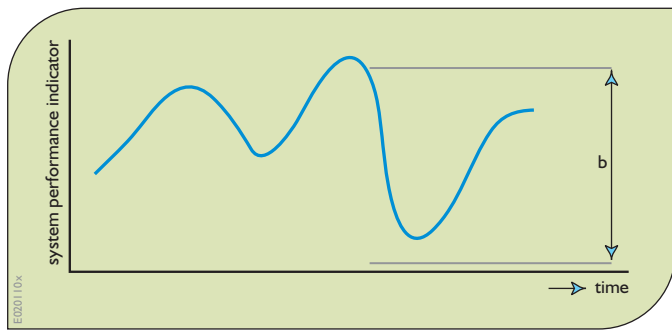
Figure 9.5. Typical time series of model output or system performance indicator values that are the result of input data variability and possible imprecision in input data measurement, parameter value estimation, model structure and errors in model solution algorithms.

Note that the time series may even contain values outside the range 'b' defined in Figure 9.4 if the confidence level of that range is less than 100%. Confidence intervals associated with less than 100% certainty will not include every possible value that might occur.

3.2. Knowledge Uncertainty

Referring to Figure 9.3, knowledge uncertainty includes model structure and parameter value uncertainties. First we consider parameter value uncertainty including boundary condition uncertainty, and then model and solution algorithm uncertainty.

3.2.1. Parameter Value Uncertainty

A possible source of uncertainty in model output results from uncertain estimates of various model parameter values. If the model calibration procedure were repeated using different data sets, different parameter values would result. Those values would yield different simulated system behaviour and, thus, different predictions. We can call this parameter uncertainty in the predictions because it is caused by imprecise parameter values. If such parameter value imprecision were eliminated, then the prediction would always be the same and so the parameter value uncertainty in the predictions would be zero. But this does not mean that predictions would be perfectly accurate.

In addition to parameter value imprecision, uncertainty in model output can result from imprecise specification of boundary conditions. These boundary conditions may be either fixed or variable. However, because they are not being computed on the basis of the state of the system, their values can be uncertain. These uncertainties can affect the model output, especially in the vicinity of the boundary, in each time step of the simulation.

3.2.2. Model Structural and Computational Errors

Uncertainty in model output can also result from errors in the model structure compared to the real system, and approximations made by numerical methods employed in the simulation. No matter how good our parameter value estimates, our models are not perfect and there is a residual model error. Increasing model complexity in order to more closely represent the complexity of the real system may not only add to the cost of data collection, but may also introduce even more parameters, and thus even more potential sources of error in model output. It is not an easy task to judge the appropriate level of model complexity and to estimate the resulting levels of uncertainty associated with various assumptions regarding model structure and solution methods. Kuczera (1988) provides an example of a conceptual hydrological modelling exercise with daily time steps where model uncertainty dominates parameter value uncertainty.

3.3. Decision Uncertainty

Uncertainty in model predictions can result from unanticipated changes in what is being modelled. These can include changes in nature, human goals, interests, activities, demands and impacts. An example of this is the deviation from standard or published operating policies by operators of infrastructure such as canal gates, pumps and reservoirs in the field, as compared to what is specified in documents and incorporated into the water systems models. Comparing field data with model data for model calibration may yield incorrect calibrations if operating policies actually implemented in the field differ significantly from those built into the models. What do operators do in times of stress?

What humans will want to achieve in the future may not be the same as what they want today. Predictions of people's future desires are clearly sources of uncertainty.

A perfect example of this can be seen in the very flat Greater Everglades region of south Florida in the United States. Fifty years ago, folk wanted the swampy region protected from floods and drained for agricultural and urban development. Today many want just the opposite, at least where there are no human settlements. They want to return to a more natural hydrological system with more wetlands and unobstructed flows, but now for ecological restoration reasons, which were not a major concern or much appreciated some half a century ago. Once the mosquitoes return, and if the sea level continues to rise, future populations who live there may want more flood control and drainage again. Who knows? Complex and changing social and economic processes influence human activities and their demands for water resources and environmental amenities over time. Some of these processes reflect changes in local concerns, interests and activities, but population migration and many economic activities and social attitudes can also reflect changing national and international trends.

Sensitivity scenarios that include human activities can help define the effects of those activities within an area. Careful attention should be given to the development of these alternative scenarios so that they capture the real forces or stresses that the system may face. The history of systems studies are full of examples where the issues studied were overwhelmed by much larger social forces resulting from, for example, the relocation of major economic activities, an oil embargo, changes in national demand for natural resources, economic recession, an act of terrorism or even war. One thing is certain: the future will be different than the past, and no one knows just how.

Surprises

Water resources managers may also want to consider how vulnerable a system is to undesirable environmental surprises. What havoc might an introduced species like the zebra mussel invading the Great Lakes of North America have in a particular watershed? Might some introduced disease suddenly threaten key plant or animal species? Might management plans have to be restructured to address the survival of species such as salmon in the Rhine River in Europe or in the Columbia River in North America? Such uncertainties are hard to anticipate when by their nature they will truly be surprises. But surprises should be expected. Hence system flexibility and adaptability should be sought to deal with changing management demands, objectives and constraints.

4. Sensitivity and Uncertainty Analyses

An uncertainty analysis is not the same as a sensitivity analysis. An uncertainty analysis attempts to describe the entire set of possible outcomes, together with their associated probabilities of occurrence. A sensitivity analysis attempts to determine the change in model output values that results from modest changes in model input values. A sensitivity analysis thus measures the change in the model output in a localized region of the space of inputs. However, one can often use the same set of model runs for both uncertainty analyses and sensitivity analyses. It is possible to carry out a sensitivity analysis of the model around a current solution, and then use it as part of a first-order uncertainty analysis.

This discussion begins by focusing on some methods of uncertainty analysis, then reviews various ways of performing and displaying sensitivity analyses.

4.1. Uncertainty Analyses

Recall that uncertainty involves the notion of randomness. If a value of a performance indicator or performance measure, like the phosphorus concentration or the depth of water at a particular location, varies, and this variation over space and time cannot be predicted with certainty, it is called a random variable. One cannot say with certainty what the value of a random variable will be but only the likelihood or probability that it will be within some specified range of values. The probabilities of observing particular ranges of values of a random variable are described or defined by a probability distribution. There are many types of distributions and each can be expressed in several ways as presented in Chapter 7.

Suppose the random variable is denoted as X. As discussed in Chapter 7, if the observed values of this random variable can only have discrete values, then the probability

Figure 9.6. Probability distributions for a discrete or continuous random variable *X*. The area under the distributions (shaded areas in a and b) is 1, and the shaded area in c is the probability that the observed value *x* of the random variable *X* will be between *u* and *v*.
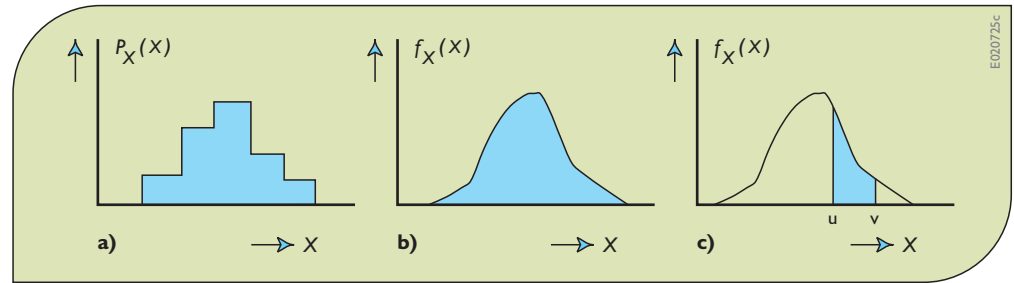
distribution of *X* is easily described by a histogram, as shown in Figure 9.6a. The sum of the probabilities for all possible outcomes must equal 1. If the random variable is a continuous variable that can assume any real value over a range of values, the probability distribution of *X* can be expressed as a continuous distribution, as shown in Figure 9.6b. The shaded area under the density function for the continuous distribution, is 1. The area between two values of the continuous random variable, such as between *u* and *v* in Figure 9.6c, represents the probability that the observed value *x* of the random variable value *X* will be within that range of values.

The probability distributions shown in Figure 9.6b and Figure 9.6c are called probability density functions (pdf) and are denoted by $f_X(x)$. The subscript *X* on P_X and f_X represents the random variable, and the variable *x* is some value of that random variable *X*.

Uncertainty analyses involve identifying characteristics of various probability distributions of model input and output variables, and subsequently functions of those random output variables that are performance indicators or measures. Often targets associated with these indicators or measures are themselves uncertain.

A complete uncertainty analysis would involve a comprehensive identification of all sources of uncertainty that contribute to the joint probability distributions of each input or output variable. Assume such analyses were performed for two alternative project plans, *A* and *B*, and that the resulting probability density distributions for a specified performance measure were as shown in Figure 9.7. The figure also identifies the costs of these two projects. The introduction of two performance criteria – cost and probability of exceeding a performance measure target, e.g. a pollutant concentration standard – introduces a conflict where a tradeoff must be made.
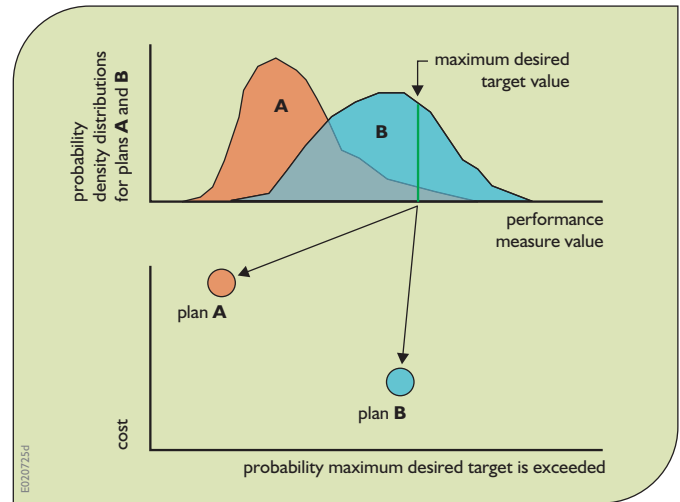
Figure 9.7. Tradeoffs involving cost and the probability that a maximum desired target value will be exceeded. In this illustration we want the lowest cost (*B* is best) and the lowest probability of exceedance (*A* is best).

4.1.1. Model and Model Parameter Uncertainties

Consider a situation as shown in Figure 9.8, in which, for a specific set of model inputs, the model outputs differ from the observed values, and for those model inputs, the observed values are always the same. Here, nothing occurs randomly. The model parameter values or model structure need to be changed. This is typically done in a model calibration process.

Given specific inputs, the outputs of deterministic models are always going to be the same each time those inputs are simulated. If for specified inputs to any simulation model the predicted output does not agree with the observed value, as shown in Figure 9.8, this could result from imprecision in the measurement of observed data. It could also result from imprecision in the model parameter values, the model structure or the algorithm used to solve the model.

Figure 9.8. A deterministic system and a simulation model of that system needing calibration or modification in its structure. There is no randomness, only parameter value or model structure errors to be identified and corrected.

Figure 9.9. A deterministic simulation model of a 'random or stochastic' system. To produce the variability in the model output that is observed in the real syste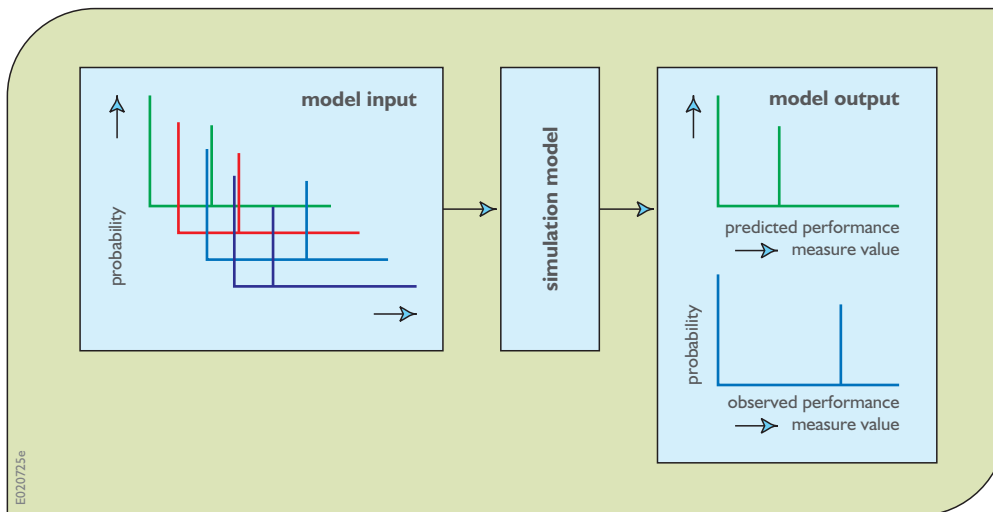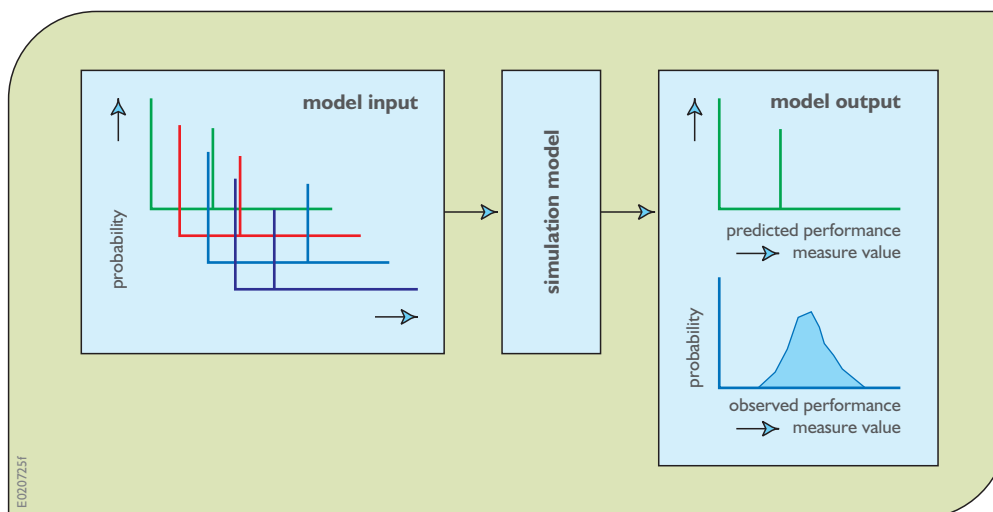m, even given the same input values, the model's parameter values may need to vary over distributions of values and/or the model structure may need modification along with additional model inputs.

Next, consider the same deterministic simulation model, but now assume at least some of the inputs are random (that is, not predictable), as may be case when random outputs of one model are used as inputs into another model. Random inputs will yield random outputs. The model input and output values can be described by probability distributions. If the uncertainty in the output is due only to the uncertainty in the input, the situation is similar to that shown in Figure 9.8. If the distribution of performance measure output values does not fit or is not identical to the distribution of observed performance measure values, then calibration of model parameter values or modification of the model structure may be needed.

If a model calibration or 'identification' exercise finds the 'best' values of the parameters to be outside reasonable ranges of values based on scientific knowledge, then the model structure or algorithm might be in error. Assuming that the algorithms used to solve the models are correct and that observed measurements of system performance vary for the same model inputs, as shown in Figure 9.9, it can be assumed that the model structure does not capture all the processes that are taking place that affect the value of the performance measures. This is often the case when relatively simple and low-resolution models are used to estimate the hydrological and ecological impacts of water and land management policies. However, even large and complex models can fail to include or adequately describe important phenomena.

In the presence of informational uncertainties, there may be considerable uncertainty about the values of the

'best' parameters during calibration. This problem becomes even more pronounced with increases in model complexity.

An example:

Consider the prediction of a pollutant concentration at some site downstream of a pollutant discharge site. Given a streamflow Q (in units of $1000 \, m^3$/day), the distance between the discharge site and the monitoring site, X(m), the pollutant decay rate constant k (day^{-1}) and the pollutant discharge W (Kg/day), we can use the following simple model to predict the concentration of the pollutant C (g/m^3 = mg/l) at the downstream monitoring site:

$C = (W/Q) \exp\{-k(X/U)\}$

In the above equation, assume the velocity U (m/day) is a known function of the streamflow Q.

In this case, the observed value of the pollutant concentration C may differ from the computed value of C even for the same inputs of W, Q, k, X and U. Furthermore, this difference varies in different time periods. This apparent variability, as illustrated in Figure 9.9, can be simulated using the same model but by assuming a distribution of values for the decay rate constant k. Alternatively, the model structure can be modified to include the impact of streamflow temperature T on the prediction of C.

$C = (W/Q) \exp\{-k\boldsymbol{\theta}^{T-20} (X/U)\}$

Now there are two model parameters, the decay rate constant k and the dimensionless temperature correction factor $\boldsymbol{\theta}$, and an additional model input, the streamflow temperature, T, in °C. It could be that the variation in streamflow temperature was the major cause of the first equation's 'uncertainty' and that the assumed parameter distribution of k was simply the result of the distribution of streamflow temperatures on the term $k\boldsymbol{\theta}^{T-20}$.

If the output were still random given constant values of all the inputs, then another source of uncertainty exists. This uncertainty might be due to additional random loadings of the pollutant, possibly from non-point sources. Once again, the model could be modified to include these additional loadings if they are knowable. Assuming these additional loadings are not known, a new random parameter could be added to the input variable W or to the right hand side of the equations above that would attempt to capture the impact on C of these additional loadings. A potential problem, however, might be the likely correlation between those additional loadings and the streamflow Q.

While adding model detail removed some 'uncertainty' in the above example, increasing model complexity will not always eliminate or reduce uncertainty in model output. Adding complexity is generally not a good idea when the increased complexity is based on processes whose parameters are difficult to measure, when the right equations are not known at the scale of application, or when the amount of data for calibration is small compared to the number of parameters.

Even if more detailed models requiring more input data and more parameter values were to be developed, the likelihood of capturing all the processes occurring in a complex system is small. Hence, those involved will have to make decisions while taking this uncertainty into account. Imprecision will always exist due to a less than complete understanding of the system and the hydrological processes being modelled. A number of studies have addressed model simplification, but only in some simple cases have statisticians been able to identify just how one might minimize modelling related errors in model output values.

The problem of determining the 'optimal' level of modelling detail is particularly important when simulating the hydrological events at many sites over large areas. Perhaps the best approach for these simulations is to establish confidence levels for alternative sets of models and then statistically compare simulation results. But even this is not a trivial or cost-free task. Increases in the temporal or spatial resolution typically require considerable data collection and/or processing, model recalibrations, and possibly the solution of stability problems resulting from the numerical methods used in the models. Obtaining and implementing alternative hydrological simulation models will typically involve considerable investment of money and time for data preparation and model calibration.

What is needed is a way to predict the variability evident in the system shown in Figure 9.9. Instead of a fixed output vector for each fixed input vector, a distribution of outputs is needed for each performance measure based on fixed inputs (Figure 9.9) or a distribution of inputs (Figure 9.10.). Furthermore, the model output distribution for each performance measure should 'match' as well
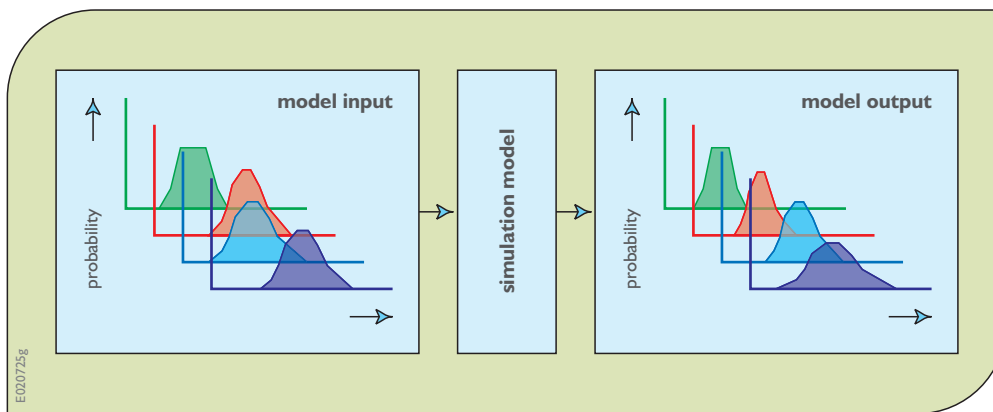
Figure 9.10. Simulating variable inputs to obtain probability distributions of predicted performance indices that match the probability distributions of observed performance values.
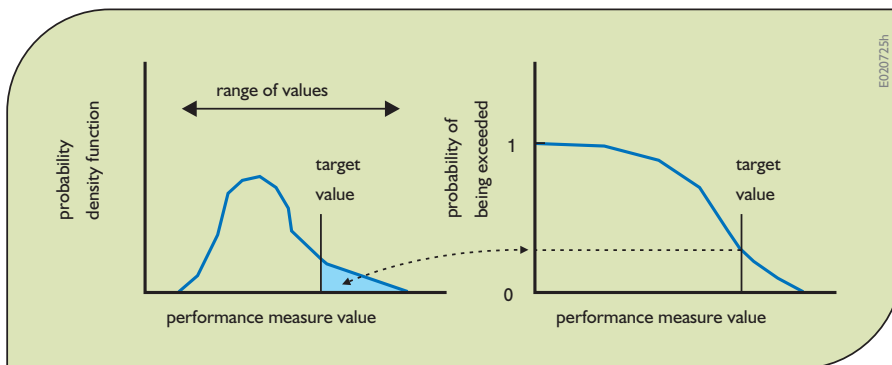

Figure 9.11. The distribution of performance measures defines a range of potential values and the likelihood that a specified target value will be exceeded. The shaded area under the density function on the left represents the probability that the target value will be exceeded. This probability is shown in the probability of exceedance plot on the right.

as possible the observed distribution of that performance measure.

4.1.2. What Uncertainty Analysis Can Provide

An uncertainty analysis takes a set of randomly chosen input values (which can include parameter values), passes them through a model (or transfer function) to obtain the distributions (or statistical measures of the distributions) of the resulting outputs. As illustrated in Figure 9.11, the output distributions can be used to

- describe the range of potential outputs of the system at some probability level
- estimate the probability that the output will exceed a specific threshold or performance measure target value.

Uncertainty analyses are often used to make general inferences, such as the following:

- estimating the mean and standard deviation of the outputs
- estimating the probability the performance measure will exceed a specific threshold

- assigning a reliability level to a function of the outputs, for example, the range of function values that is likely to occur with some probability
- describing the likelihood of different potential outputs of the system
- estimating the relative impacts of input variable uncertainties.

Implicit in any uncertainty analysis are the assumptions that statistical distributions for the input values are correct and that the model is a sufficiently realistic description of the processes taking place in the system. Neither of these assumptions is likely to be entirely correct.

4.2. Sensitivity Analyses

'Sensitivity analysis' aims to describe how much model output values are affected by changes in model input values. It is the investigation of the importance of imprecision or uncertainty in model inputs in a decision-making or modelling process. The exact character of a sensitivity analysis depends upon the particular context and the questions of concern. Sensitivity studies can provide a

general assessment of model precision when used to assess system performance for alternative scenarios, as well as detailed information addressing the relative significance of errors in various parameters. As a result, sensitivity results should be of interest to the general public, federal and state management agencies, local water resources planners and managers and model users and developers.

Clearly, upper level management and the public may be interested in more general statements of model precision, and should be provided such information along with model predictions. On the other hand, detailed studies addressing the significance and interactions among individual parameters would likely be meaningful to model developers and some model users. They can use such data to interpret model results and to identify to where their efforts to improve models and input values should be directed.

Initial sensitivity analysis studies could focus on two products:

- detailed results that guide research and assist model development efforts
- calculation of general descriptions of uncertainty associated with model predictions, so that policy decisions can reflect both the modelling efforts' best prediction of system performance and the precision of such predictions.

In the first case, knowing the relative uncertainty in model projections due to possible errors in different sets of parameters and input data should assist in efforts to improve the precision of model projections. This knowledge should also contribute to a better understanding of the relationships between model assumptions, parameters, data and model predictions.

In the second case, knowing the relative precision associated with model predictions should have a significant effect on policy development. For example, the analysis may show that, given data inadequacies, there are very large error bands associated with some model variables. When such large uncertainties exist, predictions should be used with appropriate scepticism. Incremental strategies should be explored along with monitoring so that greater experience can accumulate to resolve some of these uncertainties.

Sensitivity analysis features are available in many linear and non-linear programming (optimization) packages. They identify the changes in the values of the objective function and unknown decision-variables given a change in the model input values, and a change in levels set for various constraints (Chapter 4). Thus, sensitivity analysis can address the change in 'optimal' system performance associated with changes in various parameter values, and also how 'optimal' decisions would change with changes in resource constraint levels or target output requirements. This kind of sensitivity analysis provides estimates of how much another unit of resource would be worth, or what 'cost' a proposed change in a constraint places on the optimal solution. This information is of value to those making design decisions.

Various techniques have been developed to determine how sensitive model outputs are to changes in model inputs. Most approaches examine the effects of changes in a single parameter value or input variable assuming no changes in all the other inputs. Sensitivity analyses can be extended to examine the combined effects of multiple sources of error, as well.

Changes in particular model input values can affect model output values in different ways. It is generally true that only a relatively few input variables dominate or substantially influence the values of a particular output variable or performance indicator at a particular location and time. If the range of uncertainty of only some of the output data is of interest, then undoubtedly only those input data that significantly affect the values of those output data need be included in the sensitivity analysis.

If input data estimates are based on repeated measurements, a frequency distribution can be estimated that characterizes natural variability. The shorter the record of measurements, the greater will be the uncertainty regarding the long-term statistical characteristics of that variability. If obtaining a sufficient number of replicate measurements is not possible, subjective estimates of input data ranges and probability distributions are often made. Using a mixture of subjective estimates and actual measurements does not affect the application of various sensitivity analysis methods that can use these sets or distributions of input values, but it may affect the conclusions that can be drawn from the results of these analyses.

It would be nice to have available accurate and easy-to-use analytical methods for relating errors in input data to errors in model outputs, and to errors in system performance indicator values that are derived from model output. Such analytical methods do not exist for complex simulation models. However, methods based on

simplifying assumptions and approximations can be used to yield useful sensitivity information. Some of these are reviewed in the remainder of this chapter.

4.2.1. Sensitivity Coefficients

One measure of sensitivity is the sensitivity coefficient. This is the derivative of a model output variable with respect to an input variable or parameter. A number of sensitivity analysis methods use these coefficients. First-order and approximate first-order sensitivity analyses are two such methods that will be discussed later. Analytical methods are faced with considerable difficulties in:

- obtaining the derivatives for many models
- needing to assume mathematical (usually linear) relationships when obtaining estimates of derivatives by making small changes of input data values near their nominal or most likely values
- having large variances associated with most hydrological process models.

These have motivated the replacement of analytical methods by numerical and statistical approaches to sensitivity analysis.

Implicit in any sensitivity analysis are the assumptions that statistical distributions for the input values are correct and that the model is a sufficiently realistic description of the processes taking place in the system. Neither of these assumptions is likely to be entirely correct.

The importance of the assumption that the statistical distributions for the input values are correct is easy to check by using different distributions for the input parameters. If the outputs vary significantly, then the output is sensitive to the specification of the input distributions, and hence they should be defined with care. A relatively simple deterministic sensitivity analysis can be of value here (Benaman, 2002). A sensitivity coefficient can be used to measure the magnitude of change in an output variable Q per unit change in the magnitude of an input parameter value P from its base value P_0. Let SI_{PQ} be the sensitivity index for an output variable Q with respect to a change ΔP in the value of the input variable P from its base value P_0. Noting that the value of the output $Q(P)$ is a function of P, the sensitivity index could be defined as

$$SI_{PQ} = [Q(P_0 + \Delta P) - Q(P_0 - \Delta P)]/2\Delta P \qquad (9.1)$$

Other sensitivity indices could be defined (McCuen, 1973). Letting the index i represent a decrease and j represent an increase in the parameter value from its base value P_0, the sensitivity index SI_{PQ} for parameter P and output variable Q could be defined as

$$SI_{PQ} = \{|(Q_0 - Q_i)/(P_0 - P_i)| + |(Q_0 - Q_j)/(P_0 - P_j)|\}/2 \qquad (9.2)$$

or

$$SI_{PQ} = \max\{|(Q_0 - Q_i)/(P_0 - P_i)|, |(Q_0 - Q_j)/(P_0 - P_j)|\} \qquad (9.3)$$

A dimensionless expression of sensitivity is the elasticity index, EI_{PQ}, which measures the relative change in output Q for a relative change in input P, and could be defined as

$$EI_{PQ} = [P_0/Q(P_0)]SI_{PQ} \qquad (9.4)$$

4.2.2. A Simple Deterministic Sensitivity Analysis Procedure

This deterministic sensitivity analysis approach is very similar to those most often employed in the engineering economics literature. It is based on the idea of varying one uncertain parameter value, or set of parameter values, at a time. The ideas are applied to a water quality example to illustrate their use.

The output variable of interest can be any performance measure or indicator. Thus, one does not know if more or less of a given variable is better or worse. Perhaps too much and/or too little is undesirable. The key idea is that, whether employing physical measures or economic metrics of performance, various parameters (or sets of associated parameters) are assigned high and low values. Such ranges may reflect either the differences between the minimum and maximum values for each parameter, the 5th and 95th percentiles of a parameter's distribution, or points corresponding to some other criteria. The system model is then run with the various alternatives, one at a time, to evaluate the impact of those errors in various sets of parameter values on the output variable.

Table 9.1 illustrates the character of the results that one would obtain. Here Y_0 is the nominal value of the model output when all parameters assume the estimated best values, and $Y_{i,L}$ and $Y_{i,H}$ are the values obtained by increasing or decreasing the values of the ith set of parameters.

A simple water quality example is employed to illustrate this deterministic approach to sensitivity analysis. The analysis techniques illustrated here are just as applicable to complex models. The primary difference is that more work would be required to evaluate the various alternatives with a more complex model, and the model responses might be more complicated.

The simple water quality model is provided by Vollenweider's empirical relationship for the average phosphorus concentration in lakes (Vollenweider, 1976). He found that the phosphorus concentration, P (mg/m^3), is a function of the annual phosphorus loading rate, L (mg/m$^2 \cdot$ a), the annual hydraulic loading, q (m/a, or more exactly, m^3/m$^2 \cdot$ a), and the mean water depth, z (m):

$$P = (L/q)/[1 + (z/q)^{0.5}] \hspace{1cm} (9.5)$$

L/q and P have the same units; the denominator is an empirical factor that compensates for nutrient recycling and elimination within the aquatic lake environment.

Table 9.1. Sensitivity of model output Y to possible errors in four parameter sets containing a single parameter or a group of parameters that vary together.

Data for Lake Ontario in North America would suggest that reasonable values of the parameters are $L = 680$ mg/m^3, $q = 10.6$ m/a, and $z = 84$ m, yielding $P = 16.8$ mg/m^3. Values of phosphorus concentrations less than 10 mg/m^3 are considered oligotrophic, whereas values greater than 20 mg/m^3 generally correspond to eutrophic conditions. Reasonable ranges reflecting possible errors in the three parameters yield the values in Table 9.2.

One may want to display these results so they can be readily visualized and understood. A tornado diagram (Eschenbach, 1992) would show the lower and upper values of P obtained from variation of each parameter, with the parameter with the widest limits displayed on top, and the parameter with smallest limits on the bottom. Tornado diagrams (Figure 9.12) are easy to construct and can include a large number of parameters without becoming crowded.

An alternative to tornado diagrams is a Pareto chart showing the width of the uncertainty range associated with each variable, ordered from largest to smallest. A Pareto chart is illustrated in Figure 9.13.

Another visual presentation is a spider plot showing the impact of uncertainty in each parameter on the variable in question, all on the same graph (Eschenback, 1992; DeGarmo et al., 1993). A spider plot, Figure 9.14, shows the particular functional response of the output to each parameter on a common scale, so one needs a common metric to represent changes in all of the parameters. Here, we use percentage change from the nominal or best values.

Spider plots are a little harder to construct than tornado diagrams, and can generally include only four or five variables without becoming crowded. However, they provide a more complete view of the relationships between each parameter and the performance measure. In particular, a spider plot reveals non-linear relationships

Table 9.2. Sensitivity of estimates of phosphorus concentration (mg/m^3) to model parameter values. The two right-most values in each row correspond to the low and high values of the parameter, respectively.

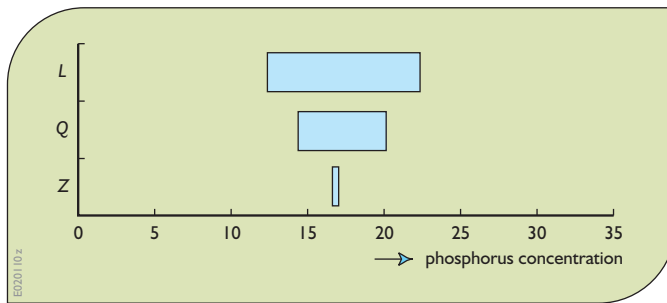|		parameter value		phosphorus concentration	
		low	high	P_{low}	P_{high}
L —	P loading (mg/m^3. a)	500	900	12.4	22.3
q —	hydraulic loading (m/a)	8	13.5	20.0	14.4
z —	mean depth (m)	81	87	17.0	16.6

Figure 9.12. A tornado diagram showing the range of the output variables representing phosphorus concentrations for high and low values of each of the parameter sets. Parameters are sorted so that the largest range is on top, and the smallest on the bottom, so the diagram looks like a tornado.
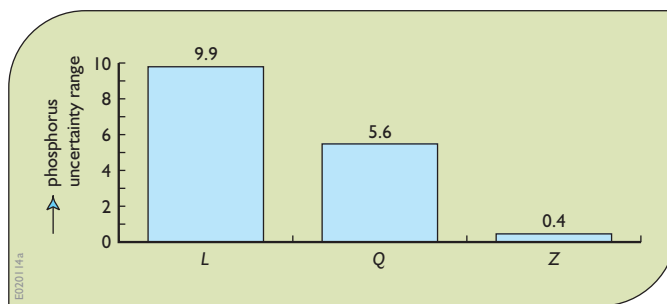

Figure 9.13. A Pareto chart showing the range of the output variable representing phosphorus concentrations resulting from high and low values of each parameter set considered.

Figure 9.14. A spider Plot illustrates the relationships between model output describing phosphorus concentrations and variations in each of the parameter sets, expressed as a percentage deviation from their nominal values.

Groups of factors

It is often the case that reasonable error scenarios would have several parameters changing together. For this reason, the alternatives have been called parameter sets. For example, possible errors in water depth would be accompanied by corresponding variations in aquatic vegetation and chemical parameters. Likewise, alternatives related to changes in model structure might be accompanied with variations in several parameters. In other cases, there may be no causal relationship among possible errors (such as model structure versus inflows at the boundary of the modelled region), but they might still interact to affect the precision of model predictions.

Combinations

If one or more non-grouped parameters interact in significant ways, then combinations of one or more errors should be investigated. However, one immediately runs into a combinatorial problem. If each of m parameters can have 3 values (high, nominal, and low), then there are 3^m combinations, as opposed to $2m + 1$ if each parameter is varied separately. (For $m = 5$, the differences are $3^5 = 243$ versus $2(5) + 1 = 11$.) These numbers can be reduced by considering instead only combinations of extremes so that only $2^m + 1$ cases need be considered ($2^5 + 1 = 33$), which is a more manageable number. However, all of the parameters would be at one extreme or the other, and such situations would be very unusual.

and the relative sensitivity of the performance measure to (percentage) changes in each variable.

In the spider plot, the linear relationship between *P* and *L* and the gentle non-linear relationship between *P* and *q* is illustrated. The range for *z* has been kept small, given the limited uncertainty associated with that parameter.

4.2.3. Multiple Errors and Interactions

An important issue that should not be ignored is the impact of simultaneous errors in more than one parameter. Probabilistic methods directly address the occurrence of simultaneous errors, but the correct joint distribution needs to be employed. With simple sensitivity analysis procedures, errors in parameters are generally investigated one at a time or in groups. The idea of considering pairs or sets of parameters is discussed here.
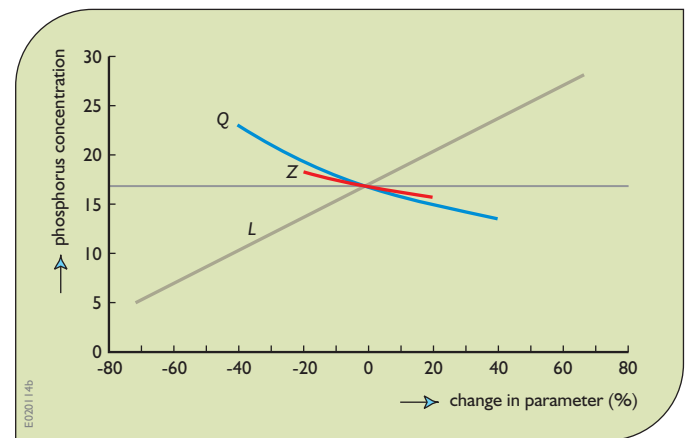
Figure 9.15. Pareto diagram showing errors in phosphorus concentrations for all combinations of pairs of input parameters errors. A '+' indicates a high value, and a '−' indicates a low value for indicated parameter. *L* is the phosphorus loading rate, *q* is the hydraulic loading, and *z* is the mean lake depth.
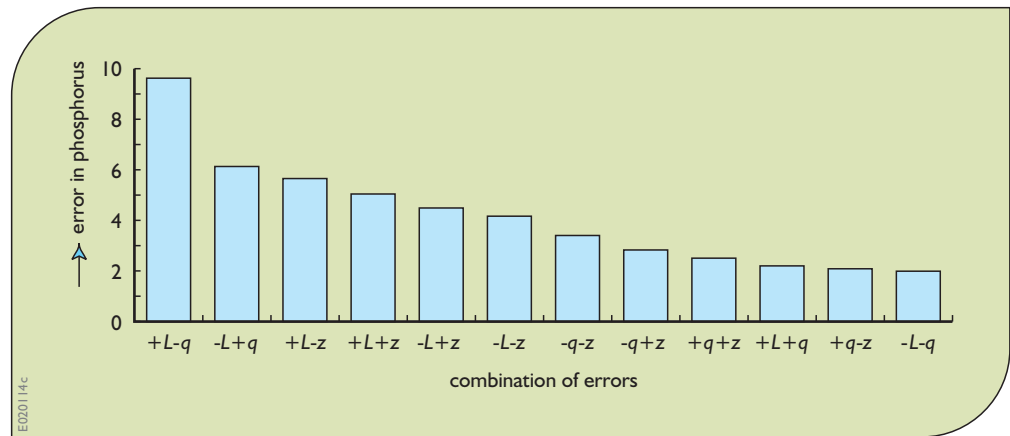
Two factors at a time

A compromise is to consider all pairs of two parameters at a time. There are $m(m-1)/2$ possible pairs of m parameters. Each parameter has a high and low value. Since there are 4 combinations of high and low values for each pair, there are a total of $2m(m-1)$ combinations. (For $m = 5$ there are 40 combinations of two parameters each having two values.)

The presentation of these results could be simplified by displaying for each case only the maximum error, which would result in $m(m-1)/2$ cases that might be displayed in a Pareto diagram. This would allow identification of those combinations of two parameters that might yield the largest errors and thus are of most concern.

For the water quality example, if one plots the absolute value of the error for all four combinations of high (+) and low (−) values for each pair of parameters, they obtain Figure 9.15.

Considering only the worst error for each pair of variables yields Figure 9.16.

Here we see, as is no surprise, that the worst error results from the most unfavourable combination of *L* and *q* values. If both parameters have their most unfavourable values, the predicted phosphorus concentration would be 27 mg/m³.

4.2.4. First-Order Sensitivity Analysis

The above deterministic analysis has trouble representing reasonable combinations of errors in several parameter sets. If the errors are independent, it is highly unlikely that any two sets would actually be at their extreme ranges at the same time. By defining probability distributions of
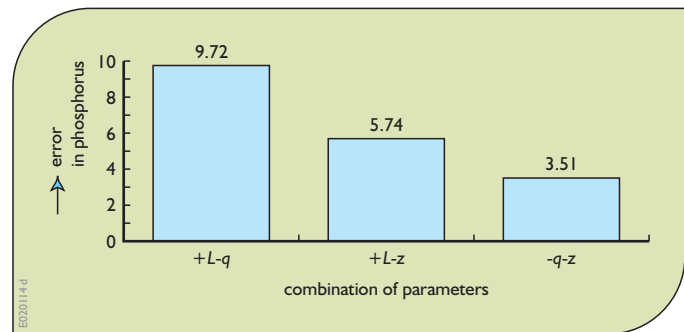
Figure 9.16. Pareto diagram showing worst error combinations for each pair of input parameters. A '+' indicates a high value, and a '−' indicates a low value for indicated parameter.

the values of the various parameter sets, and specifying their joint distributions, a probabilistic error analysis can be conducted. In particular, for a given performance indicator, one can use multivariate linear analyses to evaluate the approximate impact on the performance indices of uncertainty in various parameters. As shown below, the impact depends upon the square of the sensitivity coefficients (partial derivatives) and the variances and covariances of the parameter sets.

For a performance indicator $I = F(Y)$, which is a function $F(\cdot)$ of model outputs Y, which are in turn a function $g(P)$ of input parameters P, one can use a multivariate Taylor series approximation of F to obtain the expected value and variance of the indicator:

$$E[I] = F(\text{based on mean values of input parameters})$$
$$+ (1/2)\left\{\sum_i \sum_j [\partial F^2/\partial P_i \partial P_j]\, \text{Cov}[P_i, P_j]\right\} \qquad (9.6)$$

and

$$\mathrm{Var}[I] = \sum_i \sum_j (\partial F/\partial P_i)(\partial F/\partial P_j)\, \mathrm{Cov}[P_i, P_j] \qquad (9.7)$$

where $(\partial F/\partial P_i)$ are the partial derivatives of the function F with respect to P_i evaluated at the mean value of the input parameters P_i, and $\partial F^2/\partial P_i\partial P_j$ are the second partial derivatives. The covariance of two random input parameters P_i and P_j is the expected value of the product of differences between the values and their means:

$$\mathrm{Cov}[P_i, P_j] = E[(P_i - E[P_i])(P_j - E[P_j])] \qquad (9.8)$$

If all the parameters are independent of each other, and the second-order terms in the expression for the mean $E[I]$ are neglected, one obtains

$$E[I] = F(\text{based on mean values of input}$$
$$\text{parameters}) \qquad (9.9)$$

and

$$\mathrm{Var}[I] = \sum_i [\partial F/\partial P_i]^2\, \mathrm{Var}[P_i] \qquad (9.10)$$

(Benjamin and Cornell, 1970). Equation 9.6 for $E[I]$ shows that in the presence of substantial uncertainty, the mean of the output from non-linear systems is not simply the system output corresponding to the mean of the parameters (Gaven and Burges, 1981). This is true for any non-linear function.

Of interest in the analysis of uncertainty is the approximation for the variance $\mathrm{Var}[I]$ of indicator I. In Equation 9.10 the contribution of P_i uncertainty to the variance of I equals $\mathrm{Var}[P_i]$ times $[\partial F/\partial P_i]^2$, which are the squares of the sensitivity coefficients for indicator I with respect to each input parameter value P_i.

An Example of First-Order Sensitivity Analysis

It may appear that first-order analysis is difficult because the partial derivatives of the performance indicator I are needed with respect to the various parameters. However, reasonable approximations of these sensitivity coefficients can be obtained from the simple sensitivity analysis described in Table 9.3. In that table, three different parameter sets, P_i, are defined in which one parameter of the set is at its high value, P_{iH}, and one is at its low value, P_{iL}, to produce corresponding values (called high, I_{iH}, and low, I_{iL}) of a system performance indicator I.

		value		
		low	high	sensitivity coefficient
parameter set	1	I_{1L}	I_{1H}	$[I_{1H}-I_{1L}]/[P_{1H}-P_{1L}]$
	2	I_{2L}	I_{2H}	$[I_{2H}-I_{2L}]/[P_{2H}-P_{2L}]$
	3	I_{3L}	I_{3H}	$[I_{3H}-I_{3L}]/[P_{3H}-P_{3L}]$

Table 9.3. Approximate parameter sensitivity coefficients.

It is then necessary to estimate some representation of the variances of the various parameters with some consistent procedure. For a normal distribution, the distance between the 5th and 95th percentiles is 1.645 standard deviations on each side of the mean, or $2(1.645) = 3.3$ standard deviations. Thus, if the high/low range is thought of as approximately a 5–95 percentile range for a normally distributed variate, a reasonable approximation of the variance might be

$$\mathrm{Var}[P_i] = \{[P_{iH} - P_{iL}]/3.3\}^2 \qquad (9.11)$$

This is all that is needed. Use of these average sensitivity coefficients is very reasonable for modelling the behaviour of the system performance indicator I over the indicated ranges.

As an illustration of the method of first-order uncertainty analysis, consider the lake-quality problem described earlier. The 'system performance indicator' in this case is the model output, the phosphorus concentration P, and the input parameters, now denoted as $X = L$, q and z. The standard deviation of each parameter is assumed to be the specified range divided by 3.3. Average sensitivity coefficients $\partial P/\partial X$ were calculated. The results are reported in Table 9.4.

Assuming the parameter errors are independent:

$$\mathrm{Var}[P] = 9.18 + 2.92 + 0.02 = 12.12 \qquad (9.12)$$

The square root of 12.12 is the standard deviation and equals 3.48. This agrees well with a Monte Carlo analysis reported below.

Note that $100*(9.18/12.12)$, or about 76% of the total parameter error variance in the phosphorus concentration P, is associated in the phosphorus loading rate L and the remaining 24% is associated with the hydrological loading q. Eliminating the uncertainty in z would have a negligible impact on the overall model error. Likewise, reducing the

Table 9.4. Calculation of approximate parameter sensitivity coefficients.

variable X	units	$\partial P/\partial X$	St Dev[X]	$(\partial P/\partial X)^2$ Var[X]	%
L	mg/m^2.a	0.025	121.21	9.18	75.7
q	m/a	-1.024	1.67	2.92	24.1
z	m	-0.074	1.82	0.02	0.2

error in q would at best have a modest impact on the total error.

Due to these uncertainties, the estimated phosphorus concentration has a standard deviation of 3.48. Assuming the errors are normally distributed, and recalling that ± 1.645 standard deviations around the mean define a 5–95 percentile interval, the 5–95 percentile interval would be about

$$16.8 \pm 1.645(3.48) \text{ mg/m}^3 = 16.8 \pm 5.7 \text{ mg/m}^3$$
$$= 11.1 \text{ to } 22.5 \text{ mg/m}^3 \qquad (9.13)$$

These error bars indicate that there is substantial uncertainty associated with the phosphorus concentration P, primarily due to uncertainty in the loading rate L.

The upper bound of 22.6 mg/m^3 is considerably less than the 27 mg/m^3 that would be obtained if both L and q had their most unfavourable values. In a probabilistic analysis with independent errors, such a combination is highly unlikely.

Warning on Accuracy

First-order uncertainty analysis is indeed an approximate method based upon a linearization of the response function represented by the full simulation model. It may provide inaccurate estimates of the variance of the response variable for non-linear systems with large uncertainty in the parameters. In such cases, Monte Carlo simulation (discussed below and in Chapters 7 and 8) or the use of higher-order approximation may be required. Beck (1987) cites studies that found that Monte Carlo and first-order variances were not appreciably different, and a few studies that found specific differences. Differences are likely to arise when the distributions used for the parameters are bimodal (or otherwise unusual), or some rejection algorithm is used in the Monte Carlo analysis

to exclude some parameter combinations. Such errors can result in a distortion in the ranking of predominant sources of uncertainty. However, in most cases very similar results were obtained.

4.2.5. Fractional Factorial Design Method

An extension of first-order sensitivity analysis would be a more complete examination of the response surface using a careful statistical design. First, consider a complete factorial design. Input data are divided into discrete 'levels'. The simplest case is two levels, which can be defined as a nominal value, and a high (low) value. Simulation runs are made for all combinations of parameter levels. For n different inputs, this would require 2^n simulation runs. Hence, for a three-input variable or parameter problem, eight runs would be required. If four discrete levels of each input variable or parameter were allowed to provide a more reasonable description of a continuous variable, the three-input data problem would require 4^3 or 64 simulation runs. Clearly this is not a useful tool for large regional water resources simulation models.

A fractional factorial design involves simulating only a fraction of what is required from a full factorial design method. The loss of information prevents a complete analysis of the impacts of each input variable or parameter on the output.

To illustrate the fractional factorial design method, consider the two-level with three-input variable or parameter problem. Table 9.5 below shows the eight simulations required for a full factorial design method. The '+' and '−' symbols show the upper and lower levels of each input variable or parameter P_i where $i = 1, 2, 3$. If all eight simulations were performed, seven possible effects could be estimated. These are the individual effects of the three inputs P_1, P_2, and P_3; the three two-input variable or

simulation run	value of input	P_1	P_2	P_3	value of output - variable Y	
1		-	-	-		Y_1
2		+	-	-		Y_2
3		-	+	-		Y_3
4		+	+	-		Y_4
5		-	-	+		Y_5
6		+	-	+		Y_6
7		-	+	+		Y_7
8		+	+	+		Y_8

Table 9.5. A three-input factorial design.

parameter interactions, $(P_1)(P_2)$, $(P_1)(P_3)$, and $(P_2)(P_3)$; and the one three-input variable or parameter interaction, $(P_1)(P_2)(P_3)$.

Consider an output variable Y, where Y_j is the value of Y in the jth simulation run. Then an estimate of the effect, denoted $\delta(Y \mid P_i)$, that input variable or parameter P_i has on the output variable Y is the average of the four separate effects of varying P_i:

For $i = 1$:

$$\delta(Y \mid P_1) = 0.25[(Y_2 - Y_1) + (Y_4 - Y_3) \\ + (Y_6 - Y_5) + (Y_8 - Y_7)] \quad (9.14)$$

Each difference in parentheses is the difference between a run in which P_1 is at its upper level and a run in which P_1 is at its lower level, but the other two parameter values, P_2 and P_3, are unchanged. If the effect is equal to 0, then, in that case, P_1 has on average no impact on the output variable Y.

Similarly the effects of P_2 and P_3, on variable Y can be estimated as:

$$\delta(Y \mid P_2) = 0.25\{(Y_3 - Y_1) + (Y_4 - Y_2) + (Y_7 - Y_5) \\ + (Y_8 - Y_6)\} \quad (9.15)$$

and

$$\delta(Y \mid P_3) = 0.25\{(Y_5 - Y_1) + (Y_6 - Y_2) + (Y_7 - Y_3) \\ + (Y_8 - Y_4)\} \quad (9.16)$$

Consider next the interaction effects between P_1 and P_2. This is estimated as the average of the difference between the average P_1 effect at the upper level of P_2 and the average P_1 effect at the lower level of P_2. This is the same as the difference between the average P_2 effect at the upper level of P_1 and the average P_2 effect at the lower level of P_1:

$$\delta(Y \mid P_1, P_2) = (1/2)\{[(Y_8 - Y_7) + (Y_4 - Y_3)]/2 \\ - [(Y_2 - Y_1) + (Y_6 - Y_5)]/2\} \\ = (1/4)\{[(Y_8 - Y_6) + (Y_4 - Y_2)] \\ - [(Y_3 - Y_1) + (Y_7 - Y_5)]\} \quad (9.17)$$

Similar equations can be derived to show the interaction effects between P_1 and P_3, and between P_2 and P_3 and the interaction effects among all three inputs P_1, P_2, and P_3.

Now assume only half of the simulation runs were performed, perhaps runs 2, 3, 5 and 8 in this example. If only outputs Y_2, Y_3, Y_5, and Y_8 are available, for our example:

$$\delta(Y \mid P_3) = \delta(Y \mid P_1, P_2) \\ = 0.5\{(Y_8 - Y_3) - (Y_2 - Y_5)\} \quad (9.18)$$

The separate effects of P_3 and of P_1P_2 are not available from the output. This is the loss in information resulting from fractional instead of complete factorial design.

4.2.6. Monte Carlo Sampling Methods

The Monte Carlo method of performing sensitivity analyses, illustrated in Figure 9.17, first selects a random set of input data values drawn from their individual probability distributions. These values are then used in the simulation model to obtain some model output variable values. This process is repeated many times, each time making sure the model calibration is valid for the input data values chosen. The end result is a probability distribution of model output variables and system performance indices that results from variations and possible values of all of the input values.

Using a simple Monte Carlo analysis, values of all of the parameter sets are selected randomly from distributions describing the individual and joint uncertainty in each, and then the modelled system is simulated to obtain estimates of the selected performance indices. This must be done many times (often well over 100) to obtain a statistical description of system performance variability. The number of replications needed is generally not dependent on the number of parameters whose errors are to be analysed. One can include in the simulation the

Figure 9.17. Monte Carlo sampling and simulation procedure for finding distributions of output variable values based on distributions, for specified reliability levels, of input data values. This technique can be applied to one or more uncertain input variables at a time. The output distributions will reflect the combined effects of this input uncertainty over the specified ranges.
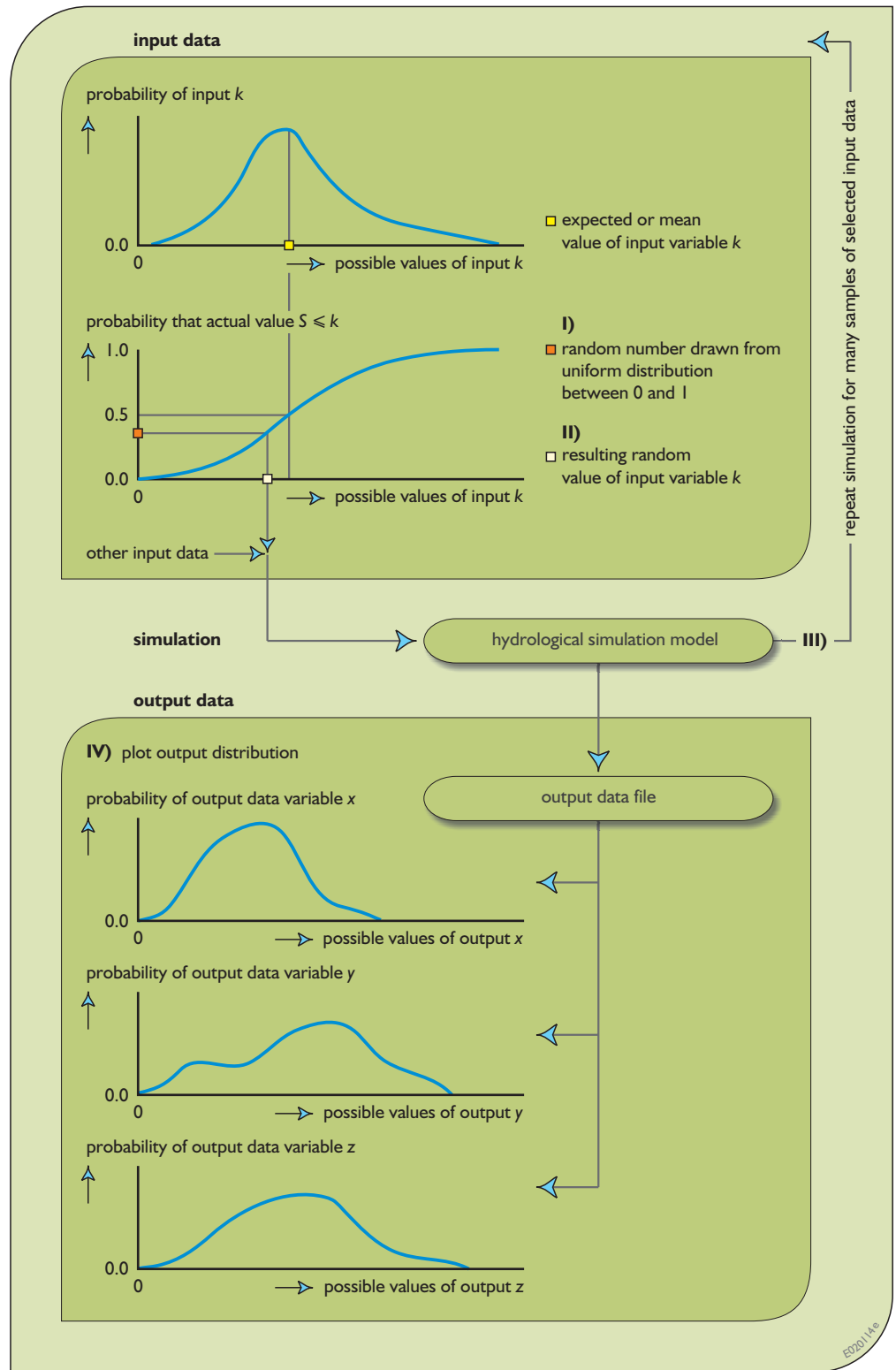
uncertainty in parameters as well as natural variability. This method can evaluate the impact of single or multiple uncertain parameters.

A significant problem that arises in such simulations is that some combinations of parameter values result in unreasonable models. For example, model performance with calibration data sets might be inconsistent with available data sets. The calibration process places interesting constraints on different sets of parameter values. Thus, such Monte Carlo experiments often contain checks that exclude combinations of parameter values that are unreasonable. In these cases the generated results are conditioned on this validity check.

Whenever sampling methods are used, one must consider possible correlations among input data values. Sampling methods can handle spatial and temporal correlations that may exist among input data values, but the existence of correlation requires defining appropriate conditional distributions.

One major limitation of applying Monte Carlo methods to estimate ranges of risk and uncertainty for model output variable values, and system performance indicator values on the basis of these output variable values, is the computing time required. To reduce the time needed to perform sensitivity analyses using sampling methods, some tricks and stratified sampling methods are available. The discussion below illustrates the idea of a simple modification (or trick) using a 'standardized' Monte Carlo analysis. The more general Latin Hypercube Sampling procedure is also discussed.

Simple Monte Carlo Sampling

To illustrate the use of Monte Carlo sampling methods, consider again Vollenweider's empirical relationship, Equation 9.5, for the average phosphorus concentration in lakes (Vollenweider, 1976). Two hundred values of each parameter were generated independently from normal distributions with the means and variances as shown in Table 9.6.

The table contains the specified means and variances for the generated values of L, q and z, and also the actual values of the means and variances of the 200 generated values of L, q, z and also of the 200 corresponding generated output phosphorus concentrations, P. Figure 9.18 displays the distribution of the generated values of P.

One can see that, given the estimated levels of uncertainty, phosphorus levels could reasonably range from below 10 to above 25. The probability of generating a value greater than 20 mg/m^3 was 12.5%. The 5–95 percentile range was 11.1 to 23.4 mg/m^3. In the figure, the cumulative probability curve is rough because only 200 values of the phosphorus concentration were generated, but these are clearly enough to give a good impression of the overall impact of the errors.

Sampling Uncertainty

In this example, the mean of the 200 generated values of the phosphorus concentration, P, was 17.07. However, a different set of random values would have generated a

Table 9.6. Monte Carlo analysis of lake phosphorus levels.

parameter	L	q	z	P
specified means and standard deviations				
mean standard deviations	680.00	10.60	84.00	—
	121.21	1.67	1.82	---
generated means and standard deviations				
mean standard deviations	674.18	10.41	84.06	17.07
	130.25	1.73	1.82	3.61

E020829s

Figure 9.18. Distribution of lake phosphorus concentrations from Monte Carlo analysis.

variable	L	q	z	P
L	1			
q	0.079	1		
z	0.130	-0.139	1	
P	0.851	-0.434	0.144	1

Table 9.7. Correlation analysis of Monte Carlo results.

	coefficient	standardized error	ratio t
intercept	18.605	1.790	10.39
L	0.025	0.000	85.36
q	-1.068	0.022	-48.54
z	-0.085	0.021	- 4.08

Table 9.8. Results of regression analysis on Monte Carlo Results.

different set of *P* values as well. Thus, it is appropriate to estimate the standard error, SE, of this average. The standard error equals the standard deviation σ of the *P* values divided by the square root of the sample size *n*:

$$SE = \sigma/(n)^{0.5} = 3.61/(200)^{0.5} = 0.25. \qquad (9.19)$$

From the central limit theorem of mathematical statistics, the average of a large number of independent values should have very nearly a normal distribution. Thus, 95% of the time, the true mean of *P* should be in the interval $17.1 \pm 1.96(0.25)$, or 16.6 to 17.6 mg/m^3. This level of uncertainty reflects the observed variability of *P* and the fact that only 200 values were generated.

Making Sense of the Results

A significant challenge with complex models is to determine from the Monte Carlo simulation which parameter errors are important. Calculating the correlation between each generated input parameter value and the output variable value is one way of doing this. As Table 9.7 below shows, on the basis of the magnitudes of the correlation coefficients, errors in *L* were most important, and those in *q* were second in importance.

One can also use regression to develop a linear model defining variations in the output on the basis of errors in the various parameters. The results are shown in the Table 9.8. The fit is very good, and $R^2 = 98\%$. If the model for *P* had been linear, a R^2 value of 100% should have resulted. All of the coefficients are significantly different from zero.

Figure 9.19. Reduction in the variance of *P* that is due to dropping from the regression model each variable individually. Clearly *L* has the biggest impact on the uncertainty in *P,* and *z* the least.

Note that the correlation between *P* and *z* was positive in Table 9.7, but the regression coefficient for *z* is negative. This occurred because there is a modest negative correlation between the generated *z* and *q* values. Use of partial correlation coefficients can also correct for such spurious correlations among input parameters.

Finally, we display a plot, Figure 9.19, based on this regression model, illustrating the reduction in the variance of *P* that results from dropping each variable

parameter	L	q	z	P
	specified means and standard deviations			
mean	680.00	10.60	84.00	—
standard deviations	121.21	1.67	1.82	---
	generated means and standard deviations			
mean	680.00	10.60	84.00	17.03
standard deviations	121.21	1.67	1.82	3.44

Table 9.9. Standardized Monte Carlo analysis of lake phosphorus levels.

variable	L	q	z	P
L	1.00			
q	0.01	1.00		
z	0.02	0.00	1.00	
P	0.85	-0.50	-0.02	1.00

Table 9.10. Correlation analysis of standardized Monte Carlo results.

individually. Clearly, L has the biggest impact on the uncertainty in P, and z the least.

Standardized Monte Carlo Analysis

Using a 'standardized' Monte Carlo analysis, one could adjust the generated values of L, q and z above so that the generated samples actually have the desired mean and variance. While making that correction, one can also shuffle their values so that the correlations among the generated values for the different parameters are near zero, as is desired. This was done for the 200 generated values to obtain the statistics shown in Table 9.9.

Repeating the correlation analysis from before (shown in Table 9.10) now yields much clearer results that are in agreement with the regression analysis. The correlation between P and both q and z are now negative, as they should be. Because the generated values of the three parameters have been adjusted to be uncorrelated, the signal from one is not confused with the signal from another.

The mean phosphorus concentration changed very little. It is now 17.0 instead of 17.1 mg/m^3.

Generalized Likelihood Estimation

Beven (1993) and Binley and Beven (1991) suggest a Generalized Likelihood Uncertainty Estimation (GLUE) technique for assessment of parameter error uncertainty using Monte Carlo simulation. It is described as a 'formal methodology for some of the subjective elements of model calibration' (Beven, 1989, p. 47). The basic idea is to begin by assigning reasonable ranges for the various parameters,

and then to draw parameter sets from those ranges using a uniform or some similar (and flat) distribution. These generated parameter sets are then used on a calibration data set so that unreasonable combinations can be rejected, while reasonable values are assigned a posterior probability based upon a likelihood measure that may reflect several dimensions and characteristics of model performance.

Let $L(P_i) > 0$ be the value of the likelihood measure assigned to the calibration sequence of the ith parameter set. Then the model predictions generated with parameter set/combination P_i are assigned posterior probability, $p(P_i)$, where

$$p(P_i) = L(P_i)/\sum_j L(P_j) \qquad (9.20)$$

These probabilities reflect the form of Bayes theorem, which is well supported by probability theory (Devore, 1991). This procedure should capture reasonably well the dependence or correlation among parameters, because *reasonable* sequences will all be assigned larger probabilities, whereas sequences that are unable to reproduce the system response over the calibration period will be rejected or assigned small probabilities.

However, in a rigorous probabilistic framework, the L would be the likelihood function for the calibration series for particular error distributions. (This could be checked with available goodness-of-fit procedures; for example, Kuczera, 1988.) When relatively ad hoc measures are adopted for the likelihood measure with little statistical validity, the $p(P_i)$ probabilities are best described as pseudo probabilities or 'likelihood' weights.

Another concern with this method is the potential efficiency. If the parameter ranges are too wide, a large

number of unreasonable or very unlikely parameter combinations will be generated. These will either be rejected or else will have small probabilities and, thus, little effect on the analysis. In this case, the associated processing would be a waste of effort. A compromise is to use some data to calibrate the model and to generate a prior or initial distribution for the parameters that is at least centred in the best range (Beven, 1993, p. 48). Use of a different calibration period to generate the $p(P_i)$ then allows an updating of those initial probabilities to reflect the information provided by the additional calibration period with the adopted likelihood measures.

After the accepted sequences are used to generate sets of predictions, the likelihood weights would be used in the calculation of means, variances and quantiles. The resulting conditional distribution of system output reflects the initial probability distributions assigned to parameters, the rejection criteria and the likelihood measure adopted to assign 'likelihood' weights.

Latin Hypercube Sampling

For the simple Monte Carlo simulations described in Chapters 7 and 8, with independent errors, a probability distribution is assumed for each input parameter or variable. In each simulation run, values of all input data are obtained from sampling those individual and independent distributions. The value generated for an input parameter or variable is usually independent of what that value was in any previous run, or what other input parameter or variable values are in the same run. This simple sampling approach can result in a clustering of parameter values, and hence both redundancy of information from repeated sampling in the same regions of a distribution and lack of information because there is no sampling in other regions of the distributions.

A stratified sampling approach ensures more even coverage of the range of input parameter or variable values with the same number of simulation runs. This can be accomplished by dividing the input parameter or variable space into sections and sampling from each section with the appropriate probability.

One such approach, Latin hypercube sampling (LHS), divides each input distribution into sections of equal probability for the specified probability distribution, and draws one observation randomly from each section. Hence the ranges of input values within each section actually occur with equal frequency in the experiment. These values from each section for each distribution are randomly assigned to those from other sections to construct sets of input values for the simulation analysis. Figure 9.20 shows the steps in constructing an LHS for six simulations involving three inputs P_j (P_1, P_2, and P_3) and six intervals of their respective normal, uniform and triangular probability distributions.

5. Performance Indicator Uncertainties

5.1. Performance Measure Target Uncertainty

Another possible source of uncertainty is the selection of performance measure target values. For example, consider a target value for a pollutant concentration based on the effect of exceeding it in an ecosystem. Which target value is best or correct? When this is not clear, there are various ways of expressing the uncertainty associated with any target value. One such method is the use of fuzzy sets (Chapter 5). Use of 'grey' numbers or intervals instead of 'white' or fixed target values is another. When some uncertainty or disagreement exists over the selection of the best target value for a particular performance measure, it seems to us the most direct and transparent way to do this is to subjectively assign a distribution over a range of possible target values. Then this subjective probability distribution can be factored into the tradeoff analysis, as outlined in Figure 9.21.

One of the challenges associated with defining and including in an analysis the uncertainty associated with a target or threshold value for a performance measure is that of communicating just what the result of such an analysis means. Referring to Figure 9.21, suppose the target value represents some maximum limit of a pollutant concentration, such as phosphorus, in the flow during a given period of time at a given site or region, and it is not certain just what that maximum limit should be. Subjectively defining the distribution of that maximum limit, and considering that uncertainty along with the
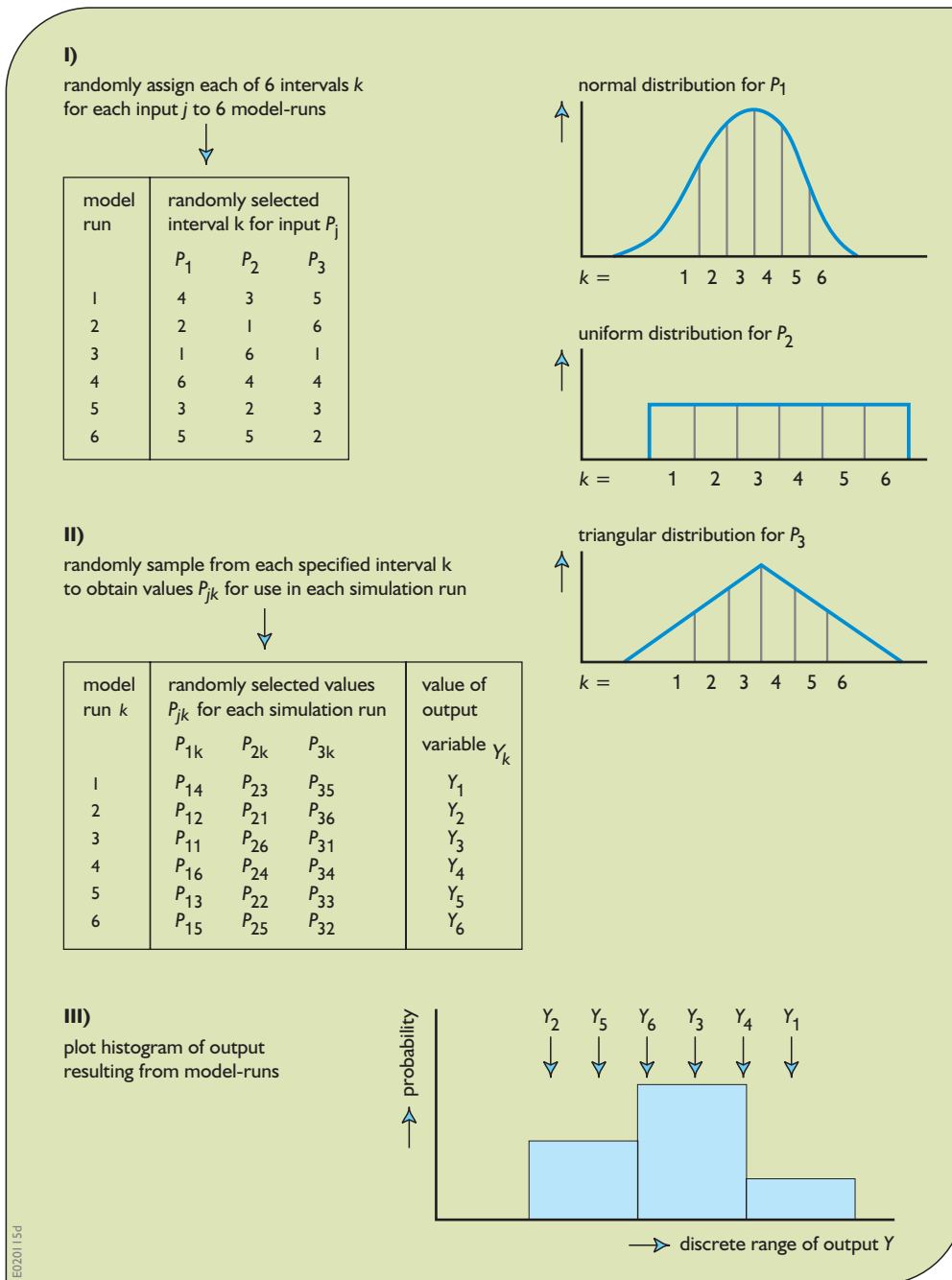
I)

randomly assign each of 6 intervals k
for each input j to 6 model-runs

model run	randomly selected interval k for input P_j		
	P_1	P_2	P_3
I	4	3	5
2	2	I	6
3	I	6	I
4	6	4	4
5	3	2	3
6	5	5	2

normal distribution for P_1

$k =$ 1 2 3 4 5 6

uniform distribution for P_2

$k =$ 1 2 3 4 5 6

II)

randomly sample from each specified interval k
to obtain values P_{jk} for use in each simulation run

model run k	randomly selected values P_{jk} for each simulation run			value of output variable Y_k
	P_{1k}	P_{2k}	P_{3k}	
I	P_{14}	P_{23}	P_{35}	Y_1
2	P_{12}	P_{21}	P_{36}	Y_2
3	P_{11}	P_{26}	P_{31}	Y_3
4	P_{16}	P_{24}	P_{34}	Y_4
5	P_{13}	P_{22}	P_{33}	Y_5
6	P_{15}	P_{25}	P_{32}	Y_6

triangular distribution for P_3

$k =$ 1 2 3 4 5 6

III)

plot histogram of output
resulting from model-runs

probability

Y_2 Y_5 Y_6 Y_3 Y_4 Y_1

discrete range of output Y

Figure 9.20. Schematic representation of a Latin hypercube sampling procedure for six simulation runs.

E020i I 5d

uncertainty (probability of exceedance function) of pollutant concentrations – the performance measure – one can attach a reliability to any probability of exceeding the maximum desired concentration value.

The 95% probability of exceedance shown on Figure 9.21, say $P_{0.95}$, should be interpreted as: 'we can be 95% confident that the probability of the maximum desired pollutant concentration being exceeded will be no greater than $P_{0.95}$.' We can be only 5% confident that the probability of exceeding the desired maximum concentration will be no greater than the lower $P_{0.05}$ value. Depending on whether the middle line through the subjective distribution of target values in Figure 9.21 represents the most likely or median target value, the associated probability of

Figure 9.21. Combining the probability distribution of performance measure values with the probability distribution of performance measure target values to estimate the confidence one has in the probability of exceeding a maximum desired target value.
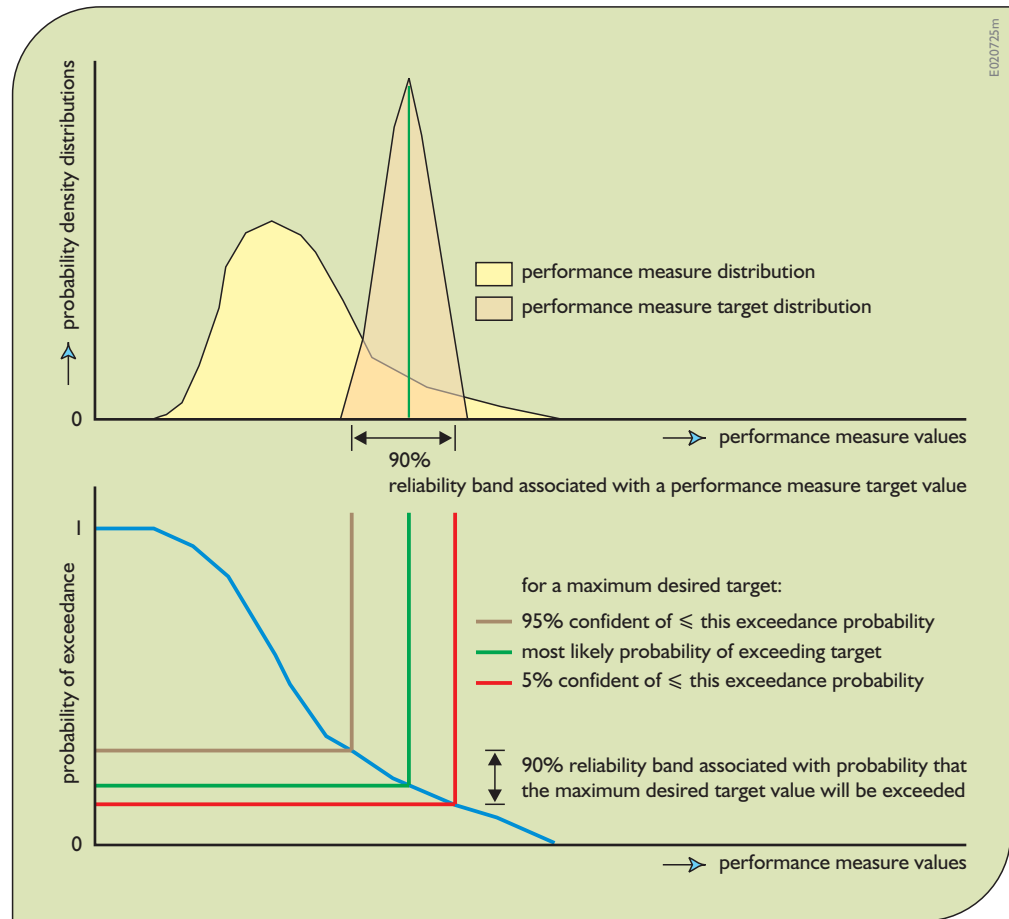
exceedance is either the most likely, as indicated in Figure 9.21, or that for which we are only 50% confident.

Figure 9.22 attempts to show how to interpret the reliabilities when the uncertain performance targets are:

- minimum acceptable levels that are to be maximized,
- maximum acceptable levels that are to be minimized, or
- optimum levels.

An example of a minimum acceptable target level might be the population of wading birds in an area. An example of a maximum acceptable target level might be, again, the phosphorus concentration of the flow in a specific wetland or lake. An example of an optimum target level might be the depth of water most suitable for selected species of aquatic vegetation during a particular period of the year.

For performance measure targets that are not expressed as minimum or maximum limits but that are the 'best' values, referring to Figure 9.22, one can state that there is a 90% reliability that the probability of

achieving the desired target is no more than β. The 90% reliability level of not achieving the desired target is at least $\alpha + \gamma$. The probability of the performance measure being too low is at least α, and the probability of the performance measure being too high is at least γ, again at the 90% reliability levels. As the reliability level decreases, the bandwidth decreases, and the probability of not meeting the target increases.

Now, clearly there is uncertainty associated with each of these uncertainty estimations, and this raises the question of how valuable is the quantification of the uncertainty of each additional component of the plan in an evaluation process. Will plan-evaluators and decision-makers benefit from this additional information, and just how much additional uncertainty information is useful?

Now consider again the tradeoffs that need to be made, as illustrated in Figure 9.7. Instead of considering a single target value, as shown on the figure, assume there is a 90% reliability or probability range associated with that single performance measure target value. Also
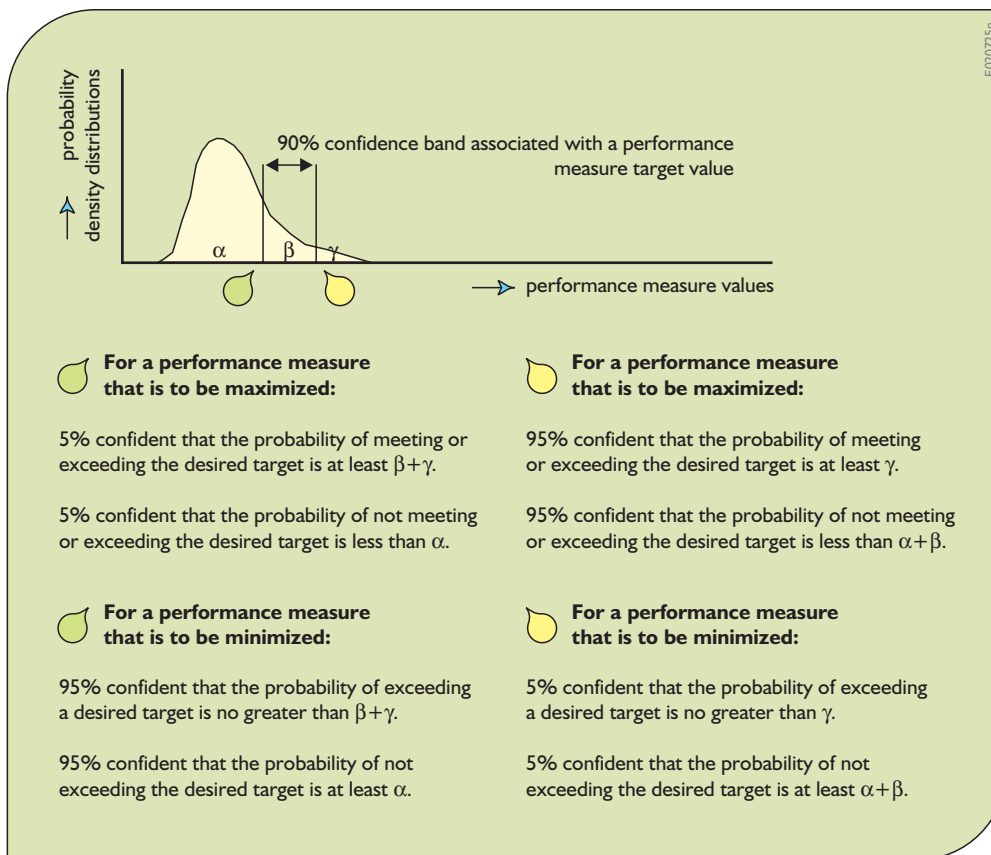
E020725n

probability density distributions ↑

90% confidence band associated with a performance measure target value

α   β   γ

performance measure values →

For a performance measure that is to be maximized:

5% confident that the probability of meeting or exceeding the desired target is at least $\beta+\gamma$.

5% confident that the probability of not meeting or exceeding the desired target is less than α.

For a performance measure that is to be maximized:

95% confident that the probability of meeting or exceeding the desired target is at least γ.

95% confident that the probability of not meeting or exceeding the desired target is less than $\alpha+\beta$.

For a performance measure that is to be minimized:

95% confident that the probability of exceeding a desired target is no greater than $\beta+\gamma$.

95% confident that the probability of not exceeding the desired target is at least α.

For a performance measure that is to be minimized:

5% confident that the probability of exceeding a desired target is no greater than γ.

5% confident that the probability of not exceeding the desired target is at least $\alpha+\beta$.

Figure 9.22. Interpreting the results of combining performance measure probabilities with performance measure target probabilities depends on the type of performance measure. The letters α, β and γ represent proportions of the probability density function of performance measure values. (Hence probabilities $\alpha + \beta + \gamma = 1$.)

assume that the target is a maximum desired upper limit, e.g. of some pollutant concentration.

In the case shown in Figure 9.23, the tradeoff is clearly between cost and reliability. In this example, no matter what reliability one chooses, Plan A is preferred to Plan B with respect to reliability, but Plan B is preferred to Plan A with respect to cost. The tradeoff is only between these two performance indicators or measures.

Consider, however, a third plan, as shown in Figure 9.24. This situation adds to the complexity of making appropriate tradeoffs. Now there are three criteria: cost, probability of exceedance and the reliabilities of those probabilities. Add to this the fact that there will be multiple performance measure targets, each expressed in terms of their maximum probabilities of exceedance and the reliability of those probabilities.

In Figure 9.24, in terms of cost the plans are ranked, from best to worst, B, C and A. In terms of the 90 reliable probability, they are ranked A, B and C, but at the 50 percent reliability level the ranking is A, C and B.

If the plan evaluation process has difficulty handling all this, it may indicate the need to focus the uncertainty analysis effort on just what is deemed important, achievable and beneficial. Then, when the number of alternatives has been narrowed down to only a few that appear to be the better ones, a more complete uncertainty analysis can be performed. There is no need and no benefit in performing sensitivity and uncertainty analyses on all possible management alternatives. Rather one can focus on those alternatives that look most promising, and then carry out additional uncertainty and sensitivity analyses only when important uncertain performance indicator values demands more scrutiny.

5.2. Distinguishing Differences Between Performance Indicator Distributions

Simulations of alternative water management infrastructure designs and operating policies require a comparison of the simulation outputs – the performance measures or indicators – associated with each alternative. A reasonable

Figure 9.23. Two plans showing ranges of probabilities, depending on the reliability, that an uncertain desired maximum (upper limit) performance target value will be exceeded. The 95% reliability levels are associated with the higher probabilities of exceeding the desired maximum target. The 5% reliability levels are associated with the more desirable lower probabilities of exceeding the desired maximum target. Plan A with reduced probabilities of exceeding the upper limit costs more than Plan B.
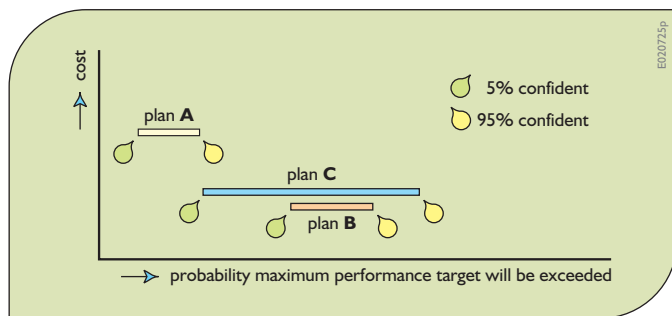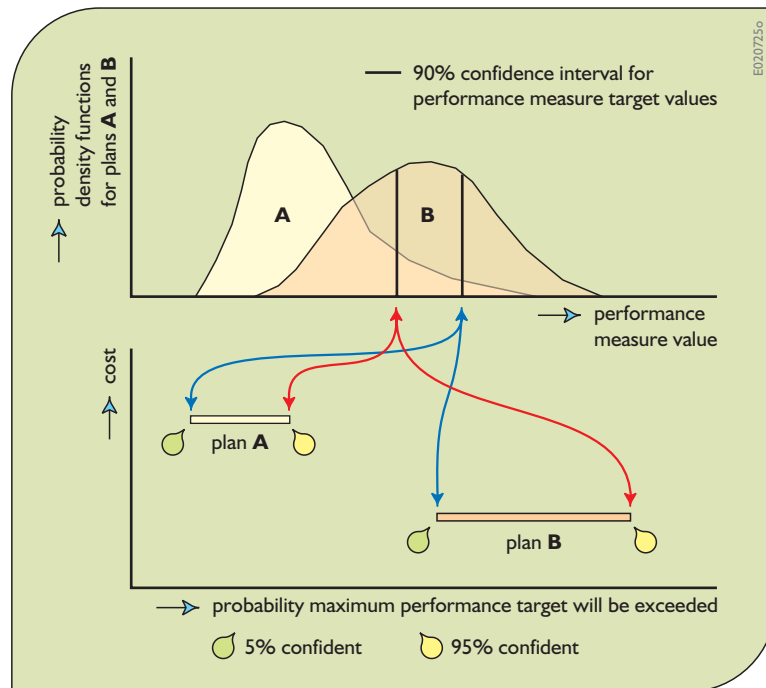

Figure 9.24. Tradeoffs among cost, probabilities and the reliability or confidence level of those probabilities. The relative ranking of plans with respect to the probability of exceeding the desired (maximum limit) target may depend on the reliability or confidence associated with that probability.

question to ask is whether the observed differences are statistically significant. Can one really tell whether one alternative is better than another, or whether the observed differences are explainable by random variations attributable to variations in the inputs and the way the system responds?

This is a common statistical issue that is addressed by standard hypothesis tests (Devore, 1991; Benjamin and Cornell, 1970). Selection of an appropriate test requires that one first resolve what type of change one expects in

the variables. To illustrate, consider the comparison of two different operating policies. Let Y_1 denote the set of output performance variable values with the first policy, and Y_2 the set of output performance variable values of the second policy. In many cases, one would expect one policy to be better than the other. One measure might be the difference in the mean of the variables; for example, $E[Y_1] < E[Y_2]$. Alternatively one could check the difference in the median (50th percentile) of the two distributions.

In addition, one could look for a change in the variability or variance, or a shift in both the mean and the variance. Changes described by a difference in the mean or median often make the most sense, and many statistical tests are available that are sensitive to such changes. For such investigations parametric and non-parametric tests for paired and unpaired data can be employed.

Consider the differences between 'paired' and 'unpaired' data. Suppose that the meteorological data for 1941 to 1990 is used to drive a simulation model generating data as described in Table 9.11.

Here there is one sample, $Y_1(1)$ through $Y_1(50)$, for Policy 1, and another sample, $Y_2(1)$ through $Y_2(50)$, for Policy 2. However, the two sets of observations are not independent. For example, if 1943 was a very dry year,

1941	Y_1 (1)	Y_2 (1)
1942	Y_1 (2)	Y_2 (2)
1943	Y_1 (3)	Y_2 (3)
1944	Y_1 (4)	Y_2 (4)
1989	Y_1 (49)	Y_2 (49)
1990	Y_1 (50)	Y_2 (50)

Table 9.11. Possible flow data from a fifty-year simulation.

then we would expect both $Y_1(3)$ for Policy 1 in that year and $Y_2(3)$ for Policy 2 to be unusually small. With such paired data, one can use a paired hypothesis test to check for differences. Paired tests are usually easier than the corresponding unpaired tests that are appropriate in other cases. For example, if one were checking for a difference in average rainfall depth between the periods 1941 to 1960 and 1961 to 1990, then one would have two sets of independent measurements for the two periods. With such data, one should use a two-sample unpaired test.

Paired tests are generally based on the differences between the two sets of output, $Y_1(i) - Y_2(i)$. These are viewed as a single independent sample. The question is then whether the differences are positive (say Y_1 tends to be larger then Y_2) or negative (Y_1 tends to be smaller), or whether positive and negative differences are equally likely (there is no difference between Y_1 and Y_2).

Both parametric and non-parametric families of statistical tests are available for paired data. The common parametric test for paired data (a one-sample t test) assumes that the mean of the differences

$$X(i) = Y_1(i) - Y_2(i) \qquad (9.21)$$

are normally distributed. The hypothesis of no difference is rejected if the t statistic is sufficiently large, given the sample size n.

Alternatively, one can employ a non-parametric test and avoid the assumption that the differences $X(i)$ are normally distributed. In such a case, one can use the Wilcoxon Signed Rank test. This non-parametric test ranks the absolute values $|X(i)|$ of the differences. If the

sum S of the ranks of the positive differences deviates sufficiently from its expected value, $n(n + 1)/4$ (were there no difference between the two distributions), one can conclude that there is a statistically significant difference between the $Y_1(i)$ and $Y_2(i)$ series. Standard statistical texts have tables of the distribution of the sum S as a function of the sample size n, and provide a good analytical approximation for $n > 20$ (for example, Devore, 1991). Both the parametric t test and the non-parametric Wilcoxon Signed Rank test require that the differences between the simulated values for each year be computed.

6. Communicating Model Output Uncertainty

Spending money on reducing uncertainty would seem preferable to spending it on ways of calculating and describing it better. Yet attention to uncertainty communication is critically important if uncertainty analyses and characterizations are to be of value in a decision-making process. In spite of considerable efforts by those involved in risk assessment and management, we know very little about how to ensure effective risk communication to gain the confidence of stakeholders, incorporate their views and knowledge, and influence favourably the acceptability of risk assessments and risk-management decisions.

The best way to communicate concepts of uncertainty may well depend on what the audiences already know about risk and the various types of probability distributions (e.g. density, cumulative, exceedance) based on objective and subjective data, and the distinction between mean or average values and the most likely values. Undoubtedly graphical representations of these ways of describing uncertainty considerably facilitate communication.

The National Research Council (NRC, 1994) addressed the extensive uncertainty and variability associated with estimating risk and concluded that risk characterizations should not be reduced to a single number or even to a range of numbers intended to portray uncertainty. Instead, the report recommended that managers and the interested public should be given risk characterizations that are both qualitative and quantitative, and both verbal and mathematical.

In some cases, communicating qualitative information about uncertainty to stakeholders and the public in general may be more effective than providing quantitative information. There are, of course, situations in which quantitative uncertainty analyses are likely to provide information that is useful in a decision-making process. How else can tradeoffs such as those illustrated in Figures 9.10 and 9.27 be identified? Quantitative uncertainty analysis can often be used as the basis of qualitative information about uncertainty, even if the quantitative information is not what is communicated to the public.

One should acknowledge to the public the widespread confusion regarding the differences between variability and uncertainty. Variability does not change through further measurement or study, although better sampling can improve our knowledge about it. Uncertainty reflects gaps in information about scientifically observable phenomena.

While it is important to communicate prediction uncertainties and the reliabilities of those uncertainties, it is equally important to clarify who or what is at risk, the possible consequences, and the severity and irreversibility of an adverse effect should a target value, for example, not be met. This qualitative information is often critical to informed decision-making. Risk and uncertainty communication is always complicated by the reliability and amounts of available relevant information as well as how that information is presented. Effective communication between people receiving information about who or what is at risk, or what might happen and just how severe and irreversible an adverse effect might be should a target value not be met, is just as important as the level of uncertainty and the reliability associated with such predictions. A two-way dialogue between those receiving such information and those providing it can help identify just what seems best for a particular audience.

Risk and uncertainty communication is a two-way activity that involves learning and teaching. Communicators dealing with uncertainty should learn about the concerns and values of their audience, their relevant knowledge and their experience with uncertainty issues. Stakeholders' knowledge of the sources and reasons for uncertainty needs to be incorporated into assessment and management and communication decisions. By listening, communicators can craft risk messages that better reflect the perspectives, technical knowledge and concerns of the audience.

Effective communication should begin before important decisions have been made. It can be facilitated in communities by citizen advisory panels, which can give planners and decision-makers a better understanding of the questions and concerns of the community, and an opportunity to test their effectiveness in communicating concepts and specific issues regarding uncertainty.

One approach to make risk more meaningful is to make risk comparisons. For example, a ten parts-per-billion target for a particular pollutant concentration is equivalent to ten seconds in over thirty-one years. If this is an average daily concentration target that is to be satisfied '99%' of the time, then this is equivalent to an expected violation of less than one day every three months.

Many perceive the reduction of risk by an order of magnitude as though it were a linear reduction. A better way to illustrate orders of magnitude of risk reduction is shown in Figure 9.25, in which a bar graph depicts better than words that a reduction in risk from one in a 1,000 (10^{-3}) to one in 10,000 (10^{-4}) is a reduction of 90% and that a further reduction to one in 100,000 (10^{-5}) is a reduction tenfold less than the first reduction of 90%. The percent of the risk that is reduced by whatever measures is a much easier concept to communicate than reductions expressed in terms of estimated absolute risk levels, such as 10^{-5}.

Risk comparisons can be helpful, but they should be used cautiously and tested if possible. There are dangers in comparing risks of diverse character, especially when the intent of the comparison is seen as reducing a risk (NRC, 1989). One difficulty in using risk comparisons is that it is not always easy to find risks that are sufficiently similar to make a comparison meaningful. How is one to compare two alternatives that have two different costs and two different risk levels, for example, as shown in Figure 9.23? One way is to perform an indifference analysis (Chapter 10), but that can lead to differing results depending who performs it. Another way is to develop utility functions using weights, where, for example reduced phosphorus load by half is equivalent to a 25% shorter hydroperiod in that area, but again each person's utility or tradeoff may differ.

At a minimum, graphical displays of uncertainty can be helpful. Consider the common system performance indicators that include:

- time-series plots for continuous time-dependent indicators (Figure 9.26 upper left),
- probability exceedance distributions for continuous indicators (Figure 9.26 upper right),

- histograms for discrete event indicators (Figure 9.26 lower left),
- overlays on maps for space-dependent discrete events (Figure 9.26 lower right).

The first three graphs in Figure 9.26 could show, in addition to the single curve or bar that represents the most likely output, a range of outcomes associated with a given confidence level. For overlays of information on maps, different colours could represent the spatial extents of events associated with different ranges of risk or uncertainty. Figure 9.27, corresponding to Figure 9.26, illustrates these approaches for displaying these ranges.

Figure 9.25. Reducing risk (x-axis) by orders of magnitude is not equivalent to linear reductions in some risk indicator (y-axis).

7. Conclusions

This chapter has provided an overview of uncertainty and sensitivity analyses in the context of hydrological or water resources systems simulation modelling. A broad range of tools are available to explore, display and quantify the sensitivity and uncertainty in predictions of key output variables and system performance indices with respect to imprecise and random model inputs and to assumptions concerning model structure. They range from relatively
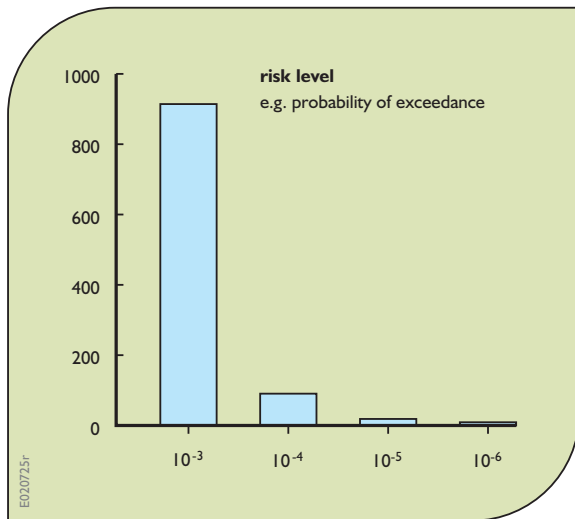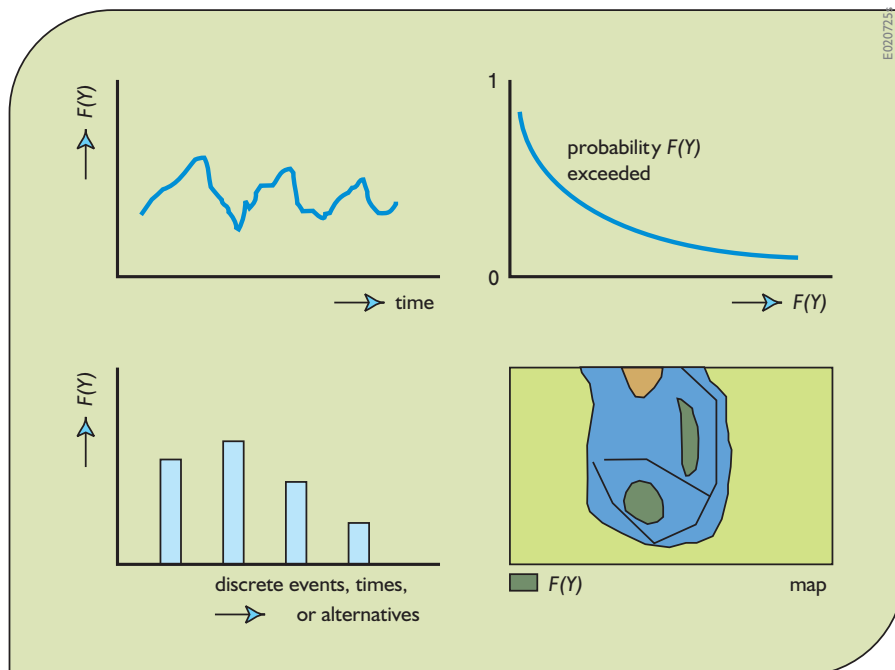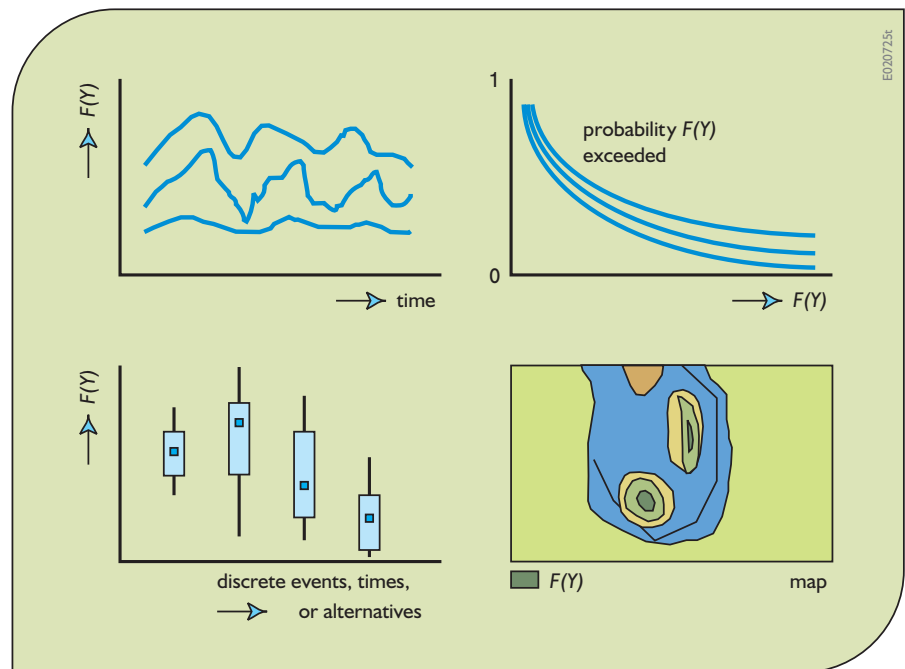

Figure 9.26. Different types of displays used to show model output Y or system performance indicator values $F(Y)$.

Figure 9.27. Plots of ranges of possible model output *Y* or system indicator values *F(Y)* for different types of displays.

simple deterministic sensitivity analysis methods to more involved first-order analyses and Monte Carlo sampling methods.

Because of the complexity of many watersheds or river basins, using Monte Carlo methods for uncertainty analyses may be a very large and unattractive undertaking. It is therefore often prudent to begin with the relatively simple deterministic procedures. These, coupled with a probabilistically based first-order uncertainty analysis method, can help quantify the uncertainty in key output variables and system performance indices, and the relative contributions of uncertainty in different input variables to the uncertainty in different output variables and system performance indices. These relative contributions may differ depending upon which output variables and indices are of interest.

A sensitivity analysis can provide a systematic assessment of the impact of parameter value imprecision on output variable values and performance indices, and of the relative contribution of errors in different parameter values to that output uncertainty. Once the key variables are identified, it should be possible to determine the extent to which parameter value uncertainty can be reduced through field investigations, development of better models and other efforts.

Model calibration procedures can be applied to individual catchments and subsystems, as well as to composite systems. Automated calibration procedures have several advantages, including the explicit use of an appropriate statistical objective function, identification of those parameters that best reproduce the calibration data set with the given objective function, and the estimations of the statistical precision of the estimated parameters.

All of these tasks together can represent a formidable effort. However, knowledge of the uncertainty associated with model predictions can be as important to management decisions and policy formulation as are the predictions themselves.

No matter how much attention is given to quantifying and reducing uncertainties in model outputs, uncertainties will remain. Professionals who analyse risk, managers and decision-makers who must manage risk, and the public who must live with risk and uncertainty, have different information needs and attitudes regarding risk and uncertainty. It is clear that information needs differ among those who model or use models, those who make substantial investment or social decisions, and those who are likely to be affected by those decisions. Meeting those needs should result in more informed decision-making,

but this comes at a cost that should be considered along with the benefits of having this sensitivity and uncertainty information.

8. References

BECK, M.B. 1987. Water quality modelling: a review of the analysis of uncertainty. *Water Resources Research*, Vol. 23, No. 8, pp. 1393–1442.

BENAMAN, J. 2002. *A systematic approach to uncertainty analysis for a distributed watershed model*. Ph.D. Thesis, School of Civil and Environmental Engineering, Cornell University, Ithaca, N.Y.

BENJAMIN, J.R. and CORNELL, C.A. 1970. *Probability, statistics and decision for civil engineers*. New York, McGraw-Hill.

BEVEN, K. 1989. Changing ideas in hydrology: the case of physically-based models. *Journal of Hydrology*. No. 105, pp. 157–72.

BEVEN, K. 1993. Prophecy, reality and uncertainty in distributed hydrological modelling. *Advances in Water Resources*, No. 16, 41–51.

BINLEY, A.M. and BEVEN, K.J. 1991. Physically-based modelling of catchment hydrology: a likelihood approach to reducing predictive uncertainty. In: D.G. Farmer and M.J. Rycroft (eds.), *Computer modelling in the environmental sciences*, Oxford, Clarendon, pp. 75–88.

DEGARMO, E.P.; SULLIVAN, W.G. and BONTADELLI, J.A. 1993. *Engineering economy*. New York, MacMillan.

DEVORE, J. 1991. *Probability and statistics for engineering and the sciences*, 3rd Edition. London, Brooks/Cole.

ESCHENBACH, T.G. 1992. Spider plots versus tornado diagrams for sensitivity analysis. *Interfaces*, No. 22, pp. 40–6.

GAVEN, D.C. and BURGES, S.J. 1981. Approximate error bounds for simulated hydrograph. *Journal of Hydraulic Engineering*, Vol. 107, No. 11, pp. 1519–34.

KUCZERA, G. 1988. On the validity of first-order prediction limits for conceptual hydrological models. *Journal of Hydrology*, No. 103, pp. 229–47.

LAL, W. 1995. *Sensitivity and uncertainty analysis of a regional model for the natural system of South Florida*. West Palm Beach, Fla., South Florida Water Management District. Draft report, November.

MCCUEN, R. 1973. The role of sensitivity analysis in hydrological modeling. *Journal of Hydrology*, Vol. 18, No. 1, pp. 37–53.

NRC (National Research Council). 1989. *Improving risk communication*. Washington, D.C., National Academy Press.

NRC (National Research Council). 1994. *Science and judgment in risk assessment*. Washington, D.C., National Academy Press.

STEDINGER, J.R. and TAYLOR, M.R. 1982. Synthetic streamflow generation, part II: Effect of parameter uncertainty. *Water Resources Research*, Vol. 18, No. 4, pp. 919–24.

VOLLENWEIDER, R.A. 1976. Advances in defining critical loading levels for phosphorus in lake eutrophication, *Memorie dell'Istituto Italiano di Idrobiologia*, No. 33, pp. 53–83.

Additional References (Further Reading)

ANDERSON, J.L. 1998. Embracing uncertainty: the interface of Bayesian statistics and cognitive psychology. *Conservation Ecology* (online), Vol. 2, No. 1, p. 2. Available from the Internet. URL: http://www.consecol.org/vol2/iss1/art2

BERRY, D.A. 1996. *Statistics: a Bayesian perspective*. Belmont, Calif., Duxbury.

CHAPRA, S.C. and RECKHOW, K.H. 1979. Expressing the phosphorus loading concept in probabilistic terms. *Journal of the Fisheries Research Board of Canada*, No. 36, pp. 225–9.

COLWELL, R.K. 1974. Predictability, consistency, and contingency of periodic phenomena. *Ecology*, Vol. 55, pp. 1148–53.

COVELLO, V.T. and MERKHOFER, M.W. 1993. *Risk assessment methods*. London, Plenum.

COVELLO, V.T. 1987. Decision analysis and risk management decision-making: issues and methods. *Risk Analysis*, No. 7, pp. 131–9.

DEUTSCH, C.V. and JOURNEL, A.G. 1992. *GS-LIB: geostatistical software library and user's guide*. New York, Oxford University Press.

DEVORE, J.L. and PECK, R. 1996. *Statistics: the exploration and analysis of data*, 3rd edn. London, Brooks/Cole.

DILKS, D.W.; CANALE, R.P. and MEIER, P.G. 1992. Development of Bayesian Monte Carlo techniques for water quality model uncertainty. *Ecological Modelling*, Vol. 62, pp. 149–62.

DILKS, D.W. and JAMES, R.T. 2002. *Application of Bayesian Monte Carlo analysis to determine the uncertainty in the Lake Okeechobee water quality model*. Proceedings of Watershed 2002. Alexandria, Va., Water Environment Federation.

DOE (Department of Energy). 1998. *Screening assessment and requirements for a comprehensive assessment*. DOE/RL-96-16, Rev. 1. Richland, Washington, US Department of Energy.

DUAN, Q.; SOROOSIAN, S. and IBBITT, R.P. 1988. A maximum likelihood criterion for use with data collected at unequal time intervals. *Water Resources Research*, Vol. 24, No. 7, pp. 1163–73.

FEDRA, K.A. 1983. Monte Carlo approach to estimation and prediction. In: M.B. Beck and G. Van Straten (eds.), *Uncertainty and forecasting of water quality*, Berlin, Springer-Verlag, pp. 259–91.

FITZ, H.C.; VOINOV, A. and COSTANZA, R. 1995. *The Everglades landscape model: multiscale sensitivity analysis*. West Palm Beach, Fla., South Florida Water Management District, Everglades Systems Research Division.

FONTAINE, T.A. and JACOMINO, V.M.F. 1997. Sensitivity analysis of simulated contaminated sediment transport. *Journal of the American Water Resources Association*, Vol. 33, No. 2, pp. 313–26.

FREY, H.C. and PATIL, S.R. 2002. Identification and review of sensitivity analysis methods. *Risk Analysis*, Vol. 22, No. 3, pp. 553–78.

GARDNER, R.H.; ROJDER, B. and BERGSTROM, U. 1983. *PRISM: a systematic method for determining the effect of parameter uncertainties on model predictions*. Technical Report, Studsvik Energiteknik AB report NW-83-/555. Sweden, Nykoping.

GELMAN, A.; CARLIN, J.; STERN, H. and RUBIN, D.B. 1995. *Bayesian data analysis*. London, Chapman and Hall.

GLEICK, J. 1987. *Chaos: making a new science*. New York, Penguin.

GUPTA, V.K. and SOROOSHIAN, S. 1985a. The automatic calibration of conceptual watershed models using derivative-based optimization algorithms. *Water Resources Research*, Vol. 21, No. 4, pp. 473–85.

GUPTA, V.K. and SOROOSHIAN, S. 1985b. The relationship between data and the precision of parameter estimates of hydrological models. *Journal of Hydrology*, No. 81, pp. 57–7.

HAIMES, Y.Y. 1998. *Risk modelling, assessment and management*. New York, Wiley.

HARLIN, J. and KUNG, C.-S. 1992. Parameter Uncertainty in simulation of design of floods in Sweden. *Journal of Hydrology*, No. 137, pp. 2009–230.

HENDRICKSON, J.D.; SOROOSHIAN, S. and BRAZIL, L.E. 1988. Comparison of Newton-type and direct-search algorithms for calibration of conceptual rainfall–runoff models. *Water Resources Research*, Vol. 24, No. 5, pp. 691–700.

HOLLING, C.S. 1978. *Adaptive environmental assessment and management*. Chichester, UK, Wiley.

IBREKK, H. and MORGAN, M.G. 1987. Graphical communication of uncertain quantities to nontechnical people. *Risk Analysis*, Vol. 7, No. 4, pp. 519–29.

ISAAKS, E.H. and SRIVASTAVA, R.M. 1989. *An introduction to applied geostatistics*. New York, Oxford University Press.

JAFFE, P.R.; PANICONI, C. and WOOD, E.F. 1988. Model calibration based on random environmental fluctuations. *Journal of Environmental Engineering*, ASCE, Vol. 114, No. 5, pp. 1136–45.

JAIN, S. and LALL, U. 2001. Floods in a changing climate: does the past represent the future? *Water Resources Research*, Vol. 37, No. 12, pp. 3193–205 DEC.

JENSEN, F.V. 2001. *Bayesian networks and decision graphs*. New York: Springer.

KANN, A. and WEYANT, J.P. 1999. A comparison of approaches for performing uncertainty analysis in integrated assessment models. *Journal of Environmental Management and Assessment*, Vol. 5, No. 1, pp. 29–46.

KELLY, E. and CAMPBELL, K. 2000. Separating variability and uncertainty: making choices. *Human and Ecological Risk Assessment, An International Journal*, Vol. 6, No. 1, February, pp. 1–13.

KELLY, E. and ROY-HARRISON, W. 1998. A mathematical construct for ecological risk: a useful framework for assessments. *Human and Ecological Risk Assessment, An International Journal*, Vol. 4, No. 2, pp. 229–41.

KELLY, E.J.; CAMPBELL, K. and HENRION, M. 1997. To separate or not to separate – that is the question: a discourse on separating variability and uncertainty in environmental risk assessments. *Learned Discourses, Society of Environmental Toxicology and Chemistry (SETAC) News*, November.

LAL, A.M.W. 2000. Numerical errors in groundwater and overland flow models. *Water Resources Research*, Vol. 36, No. 5, May, pp. 1237–48.

LAL, W.; OBEYSEKERA, J. and VAN ZEE, R. 1997. Sensitivity and uncertainty analysis of a regional simulation model for the natural system in South Florida. In: *Managing Water: Coping with Scarcity and Abundance*. Proceedings of the 27th Congress of International Association for Hydraulic Research, San Francisco, Calif., August.

LEMONS, J. (ed.). 1996. *Scientific uncertainty and environmental problem solving*. Cambridge, Mass. Blackwell.

LOPEZ, A. and LOUCKS, D.P. 1999. *Uncertainty representations in water quality prediction*. Manuscript, Civil and Environmental Engineering, Cornell University, Ithaca, N.Y.

LOUCKS, D.P.; STEDINGER, J.R. and HAITH, D.A. 1981. *Water resource systems planning and analysis*. Englewood Cliffs, N.J., Prentice-Hall.

LUDWIG, D.; HILBORN, R. and WALTERS, C. 1993. Uncertainty, resource exploitation, and conservation: lessons from history. *Science*, Vol. 260, pp. 17–36.

MAJONI, H. and QUADE, E.S. 1980. Pitfalls of analysis. New York, Wiley.

MCCARTHY, J.J.; CANZIANI, O.F.; LEARY, N.A.; DOKKEN, D.J. and WHITE, K.S. (eds.). 2001. Climate change 2001: impacts, adaptation, and vulnerability.

Contribution of Working Group II to the Third Assessment Report of the Intergovernmental Panel on Climate Change. Intergovernmental Panel on Climate Change, http://www.ipcc.ch/ pub/tar/wg2/index.htm

MCCUEN, R.H. and SNYDER, W.M. 1983. *Hydrological Modelling: statistical methods and applications*. Englewood Cliffs, N.J., Prentice-Hall.

MEIXNER, T.; GUPTA, H.V.; BASTIDAS, L.A. and BALES, R.C. 1999. Sensitivity analysis using mass flux and concentration. *Hydrological Processes*, Vol. 13, pp. 2233–44.

MISER, H.J. 1980. Operations research and systems analysis. *Science*, No. 209, pp. 174–82.

MORGAN, M.G. and HENRION, M. 1990. *Uncertainty: a guide to dealing with uncertainty in quantitative risk and policy analysis*. Cambridge, UK, Cambridge University Press.

NEARING, M.A.; DEER-ASCOUGH, L. and LAFLEN, J.F. 1990. Sensitivity analysis of the WEPP hillslope profile erosion model. *Transactions of the American Society of Agricultural Engineers*, Vol. 33, No. 3, pp. 839–49.

NRC (National Research Council). 1996. *Committee on risk characterization, understanding risk: informing decision in a democratic society*. In: P.S. Stern and H.V. Fineberg (eds.), Washington, D.C., National Academy Press.

PHILLIPS, D.L. and MARKS, D.G. 1996. Spatial uncertainty analysis: propagation of interpolation errors in spatially distributed models. *Ecological Modelling*, No. 91, pp. 213–29.

PRESS, S.J. and TANUR, J.M. 2001. *The subjectivity of scientists and the Bayesian approach*. New York: Wiley.

RECKHOW, K.H. 1994. Water quality simulation modelling and uncertainty analysis for risk assessment and decision-making. *Ecological Modelling*, No. 72, pp. 1–20.

RECKHOW, K.H. 1999. Water quality prediction and probability network models. *Canadian Journal of Fisheries and Aquatic Sciences*, No. 56, pp. 1150–8.

RECKHOW, K.H. 2002. *Applications of water models: prediction uncertainty and decision-making*. Presentation at Model Uncertainty Workshop, January 2002. West Palm Beach, Fla., South Florida Water Management District.

SALTELLI, A.; CHAN, K. and SCOTT, E.M. (eds.). 2000. *Sensitivity analysis.* Chichester, UK, Wiley.

SCHWEPPE, F.C. 1973. *Uncertain dynamic systems.* Englewood Cliffs, N.J., Prentice-Hall.

SHAPIRO, H.T. 1990. The willingness to risk failure. *Science*, Vol. 250, No. 4981, p. 609.

SIMON, I.I.A. 1988. Prediction and prescription in system modelling. 15th Anniversary of IIASA, International Institute for Applied Systems Analysis. Austria, Laxenburg.

SKLAR, F.H. and HUNSAKER, C.T. 2001. The Use and uncertainties of spatial data for landscape models: an overview with examples from the Florida Everglades. In: C.T. Hunsaker, M.F. Goodchild, M.A. Friedl and T.J. Chase (eds.), *Spatial uncertainty in ecology, implications for remote sensing and GIS applications,* Chapter 2. New York, Springer.

SOROOSHIAN, S.; DUAN, Q. and GUPTA, V.K. 1993. Calibration of rainfall–runoff models: application of global optimization to the Sacramento soil moisture accounting model. *Water Resources Research*, Vol. 29, No. 4, pp. 1185–94.

SOROOSHIAN, S.; GUPTA, V.K. and FULTON, J.L. 1983. Evaluation of maximum likelihood parameter estimation techniques for conceptual rainfall–runoff models: influence of calibration data variability and length on model credibility. *Water Resources Research*, Vol. 19, No. 1, pp. 251–9.

SOUTTER, M. and MUSY, A. 1999. Global sensitivity analyses of three pesticide leaching models using a Monte-Carlo approach. *Journal of Environmental Quality*, No. 28, pp. 1290–7.

SPEAR, R.C. and HORNBERGER, G.M. 1980. Eutrophication in Peel Inlet – II. Identification of critical uncertainties via generalized sensitivity analysis. *Water Research*, No. 14, pp. 43–9.

STOKEY, E. and ZECKHAUSER, R. 1977. *A primer for policy analysis.* New York, W.W. Norton.

SUTER, G.W. II. 1993. *Ecological risk assessment.* Boca Raton, Fla., Lewis.

TATTARI, S. and BARLUND, I. 2001. The concept of sensitivity in sediment yield modelling. *Physics and Chemistry of the Earth*, Vol. 26, No. 1, pp. 27–31.

VAN GRIENSVEN, A.; FRANCOS, A. and BAUWENS, W. 2002. Sensitivity analysis and auto-calibration of an integral dynamic model for river water quality. *Water Science and Technology*, Vol. 45, No. 9, pp. 325–332.

VAN HARN ADAMS, B.A. 1998. *Parameter distributions for uncertainty propagation in water quality modelling.* Doctoral Dissertation, Duke University. Durham, N.C.

VAN STRATEN, G. 1983. Maximum likelihood estimation of parameters and uncertainty in phytoplankton models. In: M.B. Beck and G. van Straten (eds.), *Uncertainty and forecasting of water quality*, New York, Springer-Verlag, pp. 157–71.

VON WINTERFELDT, D. and EDWARDS, W. 1986. *Decision analysis and behavioural research.* Cambridge, UK, Cambridge University Press.

WARWICK, J.J. and CALE, W.G. 1987. Determining likelihood of obtaining a reliable model. *Journal of Environmental Engineering*, ASCE, Vol. 113, No. 5, pp. 1102–19.