Estimation and propagation of parametric uncertainty in environmental models

3 authors, including:

Neil Mcintyre
The University of Queensland
161 PUBLICATIONS   **4,208** CITATIONS

Some of the authors of this publication are also working on these related projects:

Project   FRMRC2 View project

Project   Groundwater recharge estimation in the Surat Basin View project

Estimation and propagation of parametric uncertainty in environmental models

Neil McIntyre, Howard Wheater and Matthew Lees

ABSTRACT

It is proposed that a numerical environmental model cannot be justified for predictive tasks without an implicit uncertainty analysis which uses reliable and transparent methods. Various methods of uncertainty-based model calibration are reviewed and demonstrated. Monte Carlo simulation of data, Generalised Likelihood Uncertainty Estimation (GLUE), the Metropolis algorithm and a set-based approach are compared using the Streeter–Phelps model of dissolved oxygen in a stream. Using idealised data, the first three of these calibration methods are shown to converge the parameter distributions to the same end result. However, in practice, when the properties of the data and model structural errors are less well defined, GLUE and the set-based approach are proposed as more versatile for the robust estimation of parametric uncertainty. Methods of propagation of parametric uncertainty are also reviewed. Rosenblueth's two-point method, first-order variance propagation, Monte Carlo sampling and set theory are applied to the Streeter–Phelps example. The methods are then shown to be equally successful in application to the example, and their relative merits for more complex modelling problems are discussed.

Key words | calibration, propagation, stochastic, simulation, uncertainty

Neil McIntyre
Howard Wheater
Matthew Lees
Department of Civil and Environmental
  Engineering,
Imperial College of Science,
  Technology and Medicine,
London
SW7 2BU, UK
Tel: +44 207 594 6019;
Fax: +44 207 823 9401;
E-mail: *n.mcintyre@ic.ac.uk, h.wheater@ic.ac.uk*

INTRODUCTION

Motivation

The demand for far-sighted, cost-efficient solutions to engineering problems has increased simultaneously with our numerical modelling expertise and the sophistication of our computers. Consequently, numerical simulation models have become an essential part of environmental and civil engineering. They are routinely used to predict the environmental impact of engineering projects, as well as the impact of natural events on our engineering achievements.

In the 1970s and 1980s, great attention was given to improving our knowledge of the underlying environmental processes which we aimed to simulate and, as this knowledge grew, the models tended to become complex (e.g. Thomann 1998). At the same time, improving environmental databases showed that even theoretically well founded models failed to accurately replicate observations (e.g. Bierman & Dolan 1986; see also Binley *et al.*

1991). The interacting processes and the unknown spatial and temporal heterogeneities are too many to be accurately modelled or observed. This has forced environmental modellers to engage more thought in the procedure of model calibration, and to be more cautious in proposing deterministic solutions. It is becoming common procedure to include confidence limits with all model results.

Producing a reliable set of confidence limits on a model result is not difficult, given ideal circumstances. For example, to fit a linear model to observations which are normally and independently distributed with constant variance requires standard regression techniques, and derived confidence limits are theoretically sound (see Berthouex & Brown 1994). However, the natural environment is very much non-linear and this biases parameter estimates (e.g. Tellinghuisen 2000). Also, data generally carry sampling and measurement errors and are often unreliable, and, no matter how well behaved the data are,

if the structure of the model is fundamentally wrong then standard regression techniques are flawed. Clearly then, extrapolation of the model into the future also complicates the analysis, as the reliability of the model under new conditions is always in question. The problem of model equifinality (Beven & Binley 1992) means that many different proposed models may appear equally adequate when compared to the data but will give significantly different results when extrapolated to new conditions.

Scope

This paper is a review of methods of uncertainty analysis in environmental modelling. This subject area has previously been reviewed elsewhere (Beck 1987; Melching 1995; Tung 1996) and the reader is directed to this literature for additional background and discussion. This paper complements those previous works by taking a demonstrative approach to the review, aiming to give insightful comparisons between the methods using simple examples and theory. As such, it is intended to be a practical guide to the available methods, and to enable and encourage the modeller to implement them with forethought, and to interpret the results properly. Notably, this review excludes methods of recursive parameter estimation (see Beck 1987) and the use of multi-objective optimisation (see Fonseca & Fleming 1995). The utility of these methods is evident when the modelling objectives are relatively well defined by observations of the environmental system (e.g. Whitehead & Hornberger 1984; Gupta *et al.* 1998). Without diminishing the importance of these (and other omitted) methods, this paper is principally concerned with methods most used for analysis of systems for which supporting observations are relatively sparse.

The sources of uncertainty and their representation in the model

A definition of uncertainty analysis is 'the means of calculating and representing the certainty with which the model results represent reality'. The difference between a deterministic model result and reality will arise from:

- model parameter error,
- model structure error (where the model structure is the set of numerical equations which define the uncalibrated model),
- numerical errors–truncation errors, rounding errors and typographical mistakes in the numerical implementation,
- boundary condition uncertainties.

As reality can only be approximated by field data, data error analysis is a fundamental part of the uncertainty analysis. Data errors arise from:

- sampling errors (i.e. the data not representing the required spatial and temporal averages),
- measurement errors (e.g. due to methods of handling and laboratory analysis),
- human reliability.

Realising that an error-free model would equate to the error-free observations, the relationship between the actual model result M and the actual observations O can be summarised by

$$M - \varepsilon_1 - \varepsilon_2 - \varepsilon_3 - \varepsilon_4 = O - \varepsilon_5 - \varepsilon_6 - \varepsilon_7 \tag{1}$$

where ε_1–ε_4 represent the four sources of model error in the order listed above, and ε_5–ε_7 represent the three sources of data error listed above.

It is the goal of the modeller to achieve, to within an arbitrary tolerance, an error-free model by removal of ε_1–ε_4. However, the modeller is generally neither in control of model structure errors ε_2, nor numerical errors ε_3, nor boundary condition errors ε_4. Commonly, only the values of the model parameters are under the direct control of the modeller. The aim would then become one of compensating as far as possible for ε_2–ε_4 by identification of the optimum *effective* parameter values. Central to this paper is the argument that there is always some ambiguity in the 'optimum' effective parameter values caused by the unknown natures of, and inseparability of, ε_2–ε_7, and that this ambiguity can be represented by parametric uncertainty. As such, the model parameters are used as error-handling variables and are identified according to their ability to mathematically explain ε_2–ε_7. In most environmental modelling problems, significant bias in one

or more of these errors will inevitably lead to biased parameter estimates. While the ideal solution would be to eliminate bias, for example by compensatory adjustments to data or by model structure refinement, such measures are often not practical and never comprehensive. In recognition of this, the potential importance of biased model calibration will be illustrated in this paper, and significant attention is given to methods of uncertainty analysis which aim to deliver some robustness to bias.

The difficult task of identifying parameter uncertainty is generally approached using methods of calibration which derive, from the pre-calibration (*a priori*) parameter distributions, calibrated (*a posteriori*) distributions (where, for now, 'distribution' is used in the most general sense). Due to a lack of prior knowledge, the *a priori* distributions are often taken as uniform and independent (e.g. Hornberger & Spear 1980). On the other hand, the *a posteriori* distributions, constrained by the data, may be multi-modal and non-linearly interdependent (Sorooshian & Gupta 1995). Interdependency arises when the model result is simultaneously significantly affected by two or more parameters, such that the distribution of each parameter must be regarded as conditional on the value of all interdependent parameters. Therefore, it is necessary to refer to the *joint* parameter distribution, which is defined by a continuous function of *all* the parameters, and to sampled parameter *sets* rather than individual parameter values.

Model identifiability

Model identifiability is the extent to which the single most appropriate model can be identified by the modeller. It includes both model structure identifiability (i.e. that of the uncalibrated equations which form the model) and parameter identifiability (i.e. that of the parameter values given one model structure). If a set of field data can be explained equally well by two or more feasible models, then the chosen model is poorly identifiable from the data. As the alternative feasible models will not give identical results under changed boundary conditions, any predictive results will have implicit uncertainty. Poor identifiability has been generally identified as a ubiquitous issue

in environmental modelling (hence the need for uncertainty analysis), and is exacerbated with increasing model complexity (e.g. Wheater *et al.* 1986).

Parameter non-identifiability is caused when the model has too many interdependent parameters and not enough high precision data are available. The problem can be tackled by collecting additional data and/or by reducing the number of interdependent parameters. Berthouex & Brown (1994) observe 'well-designed experiments will yield precise uncorrelated parameter estimates with elliptical joint confidence regions'. This implies that the data should provide unambiguous information about every parameter as an independent entity, but in most field experiments this cannot be achieved. Alternatively the quote implies that the model should be reduced to a number of simpler, independent models. While this makes the uncertainty analysis straightforward, it may not encapsulate all the interdependent processes which the predictive model requires, and may not give insightful results. Hence there is need for a compromise between complexity and identifiability, consistent with the available data and model objective.

APPROACHES TO UNCERTAINTY-BASED MODEL CALIBRATION

Calibration is the process of tuning the model by optimisation of the set of model parameters. In traditional deterministic modelling, a single optimum parameter set is found such that model results fit the data as closely as possible. A variety of automated optimisation procedures are used (see Sorooshian & Gupta 1995). The closeness of fit is quantified by one or more objective functions (OFs). The OF is often some expression of the sum of the squared residuals (of the data and the model result) (see Weglarczyk 1998). However, the OF should be designed according to the nature of both the data errors (Sorooshian & Dracup 1980; Valdes *et al.* 1980) and the model errors (Beven & Binley 1992), as the optimum parameter values depend intimately on both. In an uncertainty-based calibration, where it is recognised that use of one optimum parameter set will give results of

limited insight, the modeller is interested in the response of the OF over the entire *a priori* range of parameter sets, i.e. the OF response surface (see Berthouex & Brown 1994). Analysis of the response surface is the means of deriving the calibrated parameter distributions. This discussion will describe and demonstrate different approaches to this analysis.

Objective functions and likelihood measures

The method of maximum likelihood (see Ang & Tang 1975) is the basis for traditional calibration, and it is a necessary starting point for this discussion. If the OF is defined as a likelihood estimator of the model then for each trial model

$$\text{OF} = P(\varepsilon_1)P(\varepsilon_2|\varepsilon_1)\prod_i P[\varepsilon_i|(\varepsilon_1 \cap \varepsilon_2 \cap \dots \cap \varepsilon_{i-1}];$$
$$i = 3, 4, \dots, N \qquad (2)$$

where ε_i is the *i*th of N model residuals (i.e. the difference between the *i*th of N available data points and the corresponding model result), $P(\varepsilon_i)$ is the probability density of ε_i, and $P(\varepsilon_2 \mid \varepsilon_1)$ signifies the probability of ε_2 assuming ε_1 has already happened. If the N residuals are assumed to be independent and normally distributed with zero mean and constant variance σ^2, and there are F degrees of freedom (i.e. parameters to be calibrated), then Equation (2) reduces to

$$\text{OF} = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp(-0.5(N-F)). \qquad (3)$$

If N and F are constant, which is likely during a model calibration, Equation (3) is reduced to

$$\text{OF} = \frac{K}{(\sigma^2)^{N/2}} \qquad (4)$$

where K is a constant. Therefore, assuming that the sum of the squared residuals divided by an arbitrary constant is an unbiased estimator of σ^2, the least sum of squared residuals maximises the likelihood (Box & Jenkins 1970). Usually, one or more of the assumptions used in the derivation of Equation (4) is not valid. If more than one

system response is being modelled and monitored then σ^2 cannot be taken as constant, and Equation (4) becomes

$$\text{OF} = \prod_{r=1}^{R} \frac{K}{(\sigma^2_r)^{N/2}} \qquad (5)$$

where R is the number of responses each measured at N locations (in time and/or space). For finding the maximum likelihood, this is equivalent to minimisation of the sum of weighted squared residuals, assuming the responses are independent. Similarly, if the variance changes in time and/or space with the magnitude of the response then an appropriate weighting scheme may be used (e.g. Sorooshian & Dracup 1980). For autocorrelated residuals, Romanowicz *et al.* (1994) describe a suitable likelihood estimator.

The OFs in Equations (2)–(5) give the probability density of the data sample (say, data sample k) occurring, given the model result. If this model result is defined by a set of parameters (α_i) sampled from the *a priori* joint parameter distribution and applied to a chosen model structure (model structure$_j$), then

$$P[\text{data sample}_k \mid (\alpha_i \mid \text{model structure}_j)] = \text{OF}_i. \qquad (6)$$

This conditional probability may be manipulated using Bayes' theorem (see Ang & Tang 1975) to give the probability of α_i given the chosen model structure and given the data sample:

$$P[(\alpha_i|\text{model structure}_j)|\text{data sample}_k] \times$$
$$\frac{P(\text{data sample}_k)}{P[(\alpha_i|\text{model structure}_j)} = \text{OF}_i. \qquad (7)$$

If only one data sample is considered, with no explicit attention to data sampling error, then $P(\text{data sample}_k) = 1$. Furthermore, if it is considered that all of S sampled parameter sets have equal *a priori* probability then $P(\alpha_i \mid \text{model structure}_j)$ is equal to $1/S$. In summary

$$P[(\alpha_i|\text{model structure}_j)|\text{data sample}_k] = \frac{\text{OF}_i}{S}. \qquad (8)$$

The standardised likelihood (so that all the discrete OFs total unity), P', can be regarded as a point estimate of

probability mass from the *a posteriori* joint parameter distribution:

$$P'[(\alpha_i|\text{model structure}_j)|\text{data sample}_k] = \frac{\text{OF}_i}{\sum\limits_{l=1,S}\text{OF}_l} . \qquad (9)$$

The number of sampled parameter sets, in particular the adequacy of the sampling of the important parameter interactions, is fundamental to the reliability of the results (Cochran 1977; Kuczera & Parent 1998). As the important interactions are not known *a priori*, the sampling is often randomised, which is known as Monte Carlo sampling (see Ang & Tang 1984). The limitation of this approach is the large number of parameter samples, and hence model runs, generally required to achieve convergence of the *a posteriori* parameter distribution. However, this can be mitigated by numerous variance reduction techniques, for example Latin hypercube sampling and stratified random sampling (MacKay *et al.* 1979).

Data sampling error arises from the fact that, if many data sets are sampled from the same population, no two sets will give the same likelihood measure for a given model. The data sampling error can be incorporated in the calibration using the theorem of total probability:

$$P[(\alpha_i|\text{model structure}_j] =$$

$$\sum_{k=1,D} P[(\alpha_i|\text{model structure}_j|\text{data sample}_k] \times \qquad (10)$$

$$P[\text{data sample}_k])$$

where D is the number of sampled data sets. If it is supposed that the data sampling error is the main source of uncertainty, and only the maximum likelihood model for each data realisation need be considered, then

$$P[\alpha'_k \mid \text{model structure}_j] = P[\text{data sample}_k] \qquad (11)$$

where α'_k is the maximum likelihood set of parameters set found for the kth sample of data. Thereby, many different data sets from the same population are required, usually an unrealistic requirement in the context of environmental sampling programmes. Alternatively, one available sample can be used to estimate the distributional properties of the residuals (e.g. Ang & Tang 1975). Equation (11) can then be solved by randomly simulating data on the basis of
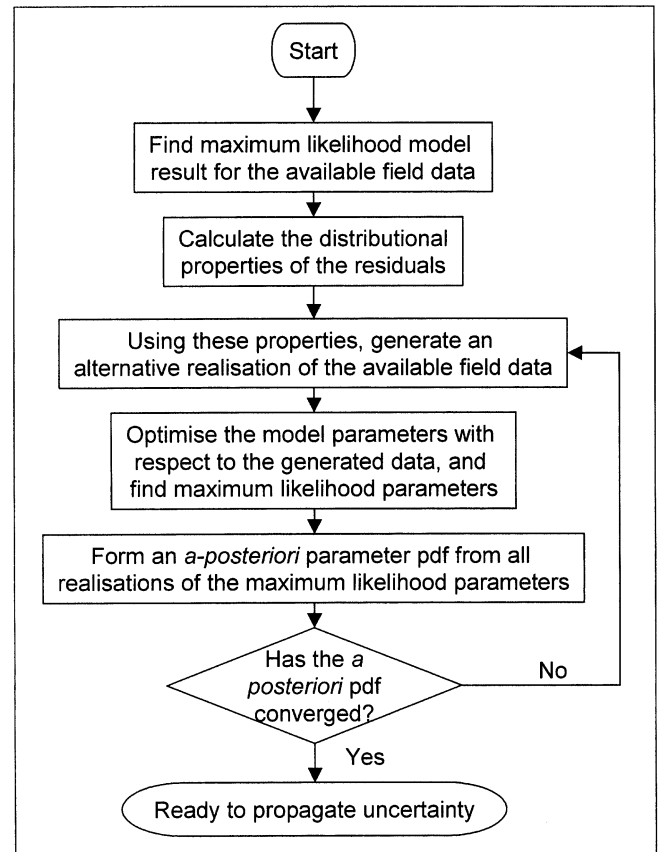

Figure 1 │ Monte Carlo calibration procedure.

these estimated properties. This is Monte Carlo simulation of data, an established basis for estimation of parameter uncertainty (Rubinstein 1981; Shao & Tu 1995). Such a procedure is shown in Figure 1. As an alternative to simulation of data, residual bootstrapping can be used (e.g. Shao 1996). This uses sub-samples of the residuals (of the available data and the maximum likelihood model result) as the kth data realisation. As with Monte Carlo simulation of data, residual bootstrapping requires initial assumptions about the maximum likelihood, but it avoids assumptions regarding the variance.

Possibility theory and the HSY method

A problem with the maximum likelihood estimator is that simplifying assumptions are required about the nature of

the data errors. In environmental monitoring, the sampling location and methods of measurement generally cause unknown biases in the maximum likelihood estimate. Various methods may be tried to improve robustness to assumptions regarding data bias. One approach is to use possibility theory (Zadeh 1978; see also Wierman 1996). A possibility distribution describes the perceived possibility of an event where the maximum possibility is 1 and the minimum is 0. In possibility theory, the rules of union and intersection differ from those in probability theory. For independent, random variables X and Y,

$$\text{Possibility}(X \cap Y) = \text{Minimum} \quad [\text{Possibility}(X), \quad \text{Possibility}(Y)] \tag{12a}$$

$$\text{Possibility}(X \cup Y) = \text{Maximum} \quad [\text{Possibility}(X), \quad \text{Possibility}(Y)] \tag{12b}$$

Applying possibility theory to model calibration requires a subjective measure of the possibility of the outcome of each candidate model. Using Equation (12a), for example, the possibility of any model result is the model residual (out of all N model residuals) perceived to be the least likely. Although the significance of the remaining $N-1$ residuals would be lost, the robustness to data bias would be increased by avoidance of the multiplicative likelihood estimator.

Another particular appeal of applying possibility theory to model calibration is that it provides a convenient basis for calibrating the model using subjectively defined support criteria. While such reasoning can be based on interpretation of data it may also be knowledge-based. That is, the possibility of any candidate model can be judged on the basis of non-numeric (even non-documented) knowledge about the environmental system rather than by 'hard' data.

Hornberger & Spear (1980) suggested a groundbreaking approach to calibration of environmental models which has distinct parallels with possibility theory. In their method, an *a priori* parameter set, applied to a given model structure, is considered to be a possible model of the system if the corresponding model result lies wholly within a set of characteristic system behaviour. The characteristic behaviour is defined by subjective reasoning

which may include analysis of available data. The result of this approach to calibration is an *a posteriori* sample of equally possible parameters and a complementary sample of impossible parameter sets. Van Straten & Keesman (1991) demonstrate how the *a posteriori* sample of possible parameters can be propagated to a range of possible results. Statistical comparison of the contents of these sets can robustly quantify model sensitivity to individual parameters (e.g. Spear & Hornberger 1980; Chen & Wheater 1999), and so the method is often referred to as Regional (or Global) Sensitivity Analysis. After Beck (1987), the method is referred to hereafter as the Hornberger–Spear–Young (HSY) algorithm and an interpretation of the algorithm is illustrated in Figure 2.

Generalised Likelihood Uncertainty Estimation (GLUE)

Beven & Binley (1992) developed the HSY method into their Generalised Likelihood Uncertainty Estimation (GLUE), so that every possible model was weighted with a likelihood. The array of likelihoods (for each model structure) is interpreted as point estimates of probability from the joint parameter distribution for that model. As new data becomes available, the distributions can be updated. The predictions from alternative model structures with their own joint parameter distributions can be combined using Bayes' method. The key feature of GLUE is that the modeller designs an OF which is taken as a measure of the likelihood. It is emphasised that the estimated uncertainty depends largely on the user's design of likelihood measure (e.g. Freer *et al.* 1996), and that the basis of the design should be explicit. In particular, the GLUE likelihood measure should not be interpreted as a statistical likelihood estimator, as used in Equations (2)–(5), unless it is specifically designed as such (e.g. Romanowicz *et al.* 1994) with regard to the data and model errors.

A major problem with the likelihood estimators of Equations (2)–(5), which may be addressed using GLUE, is the assumption that the model is correct and uncertainty only arises from the statistical significance with which the data can define the optimum solution. If the model is biased with respect to the observations (i.e. incapable of producing uncorrelated residuals), this has the unsatisfactory effect that the parameter uncertainty tends to reduce,
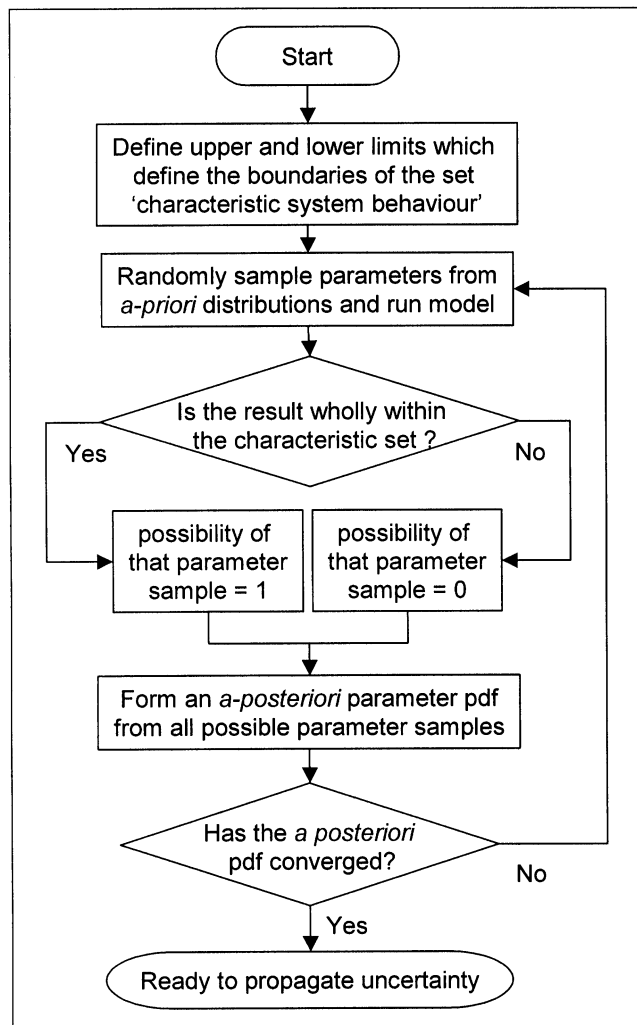
stochastic model may be judged by its ability to encompass the measurements, irrespective of the measurement errors. However, simple manipulation of likelihood estimators will not solve the problem of model or data bias, for which GLUE must be used in its more general, subjective capacity.

Adaptation of parameter distributions

It has been described how the OF evaluated for any parameter set can be interpreted as a point estimate of probability mass or possibility. Thus, using Monte Carlo simulation of the parameters, a response surface equivalent to an *a posteriori* distribution can be approximated. However, many thousands of *a priori* parameter samples may be required for an adequate approximation to be made (e.g. Kuczera & Parent 1998). To improve the efficiency of the calibration, attempts have been made to adapt the *a priori* distribution to an *a posteriori* form. Such approaches are commonly called adaptive random searches (ARSs). Types of ARS include genetic algorithms (Holland 1975), shuffled complex evolution (Duan *et al.* 1993) and Monte Carlo Markov chains (see Rutenbar 1989), all of which have proved useful for environmental model calibration and for uncertainty analysis (e.g. Mailhot *et al.* 1997; Mulligan & Brown 1998; Thyer *et al.* 1999). Careful thought is required before applying an ARS to estimation of parametric uncertainty because the achieved *a posteriori* distribution depends on the particular ARS and the convergence criteria, as well as the data, the model and the OF. So, while a distribution representing uncertainty may be derived, the significance of this uncertainty is not necessarily helpful.

Here, a Monte Carlo Markov chain model proposed by Metropolis *et al.* (1953) is described. The algorithm uses a Markov chain process (see Rutenbar 1989) which, in essence, assumes that the current state of a system dictates the probability of moving to any proposed new state. The Metropolis algorithm was originally developed to simulate the stochastic behaviour of a system of particles at thermal equilibrium. Applied to model calibration, it adapts the

Figure 2 │ HSY calibration procedure.

as high likelihood is exclusive to those parameter sets (of course, themselves biased) which compensate for the model bias. Furthermore, when the number of data points is high, parameter uncertainty is low and the associated confidence limits on the model results will not encompass many measurements of the real system. This may be solved within the GLUE framework, for example, by prescribing a value of N (Equations (3)–(5)) which is less than the number of data points, so increasing the parameter variance (e.g. Freer *et al.* 1996). This has a distinct advantage in regulation-driven modelling exercises, where the

population of parameters until the OF (in this case to be minimised) is described by the distribution

$$P(\alpha_i) = \frac{1}{K_{MET}} e^{-OF_i/A} \qquad (13)$$

where K_{MET} is a standardisation constant such that the total of all $P(\alpha_i)$ is unity, A is a case-dependent constant and α_i is the ith parameter set in the derived population. Note that, while the distribution of the accepted OFs converges to the Gaussian form of Equation (13), the distribution of the accepted parameter sets depends upon the relationship between the model and the OF. The algorithm starts from an arbitrary location in the *a priori* parameter space. From then on, the probability of any sampled parameter set α_i being accepted into the population depends entirely on comparison of OF_i with that of the last accepted set, OF_{i-1}. This probability is defined by Equations (14a) and (14b):

$$P(\alpha_{i-1} \rightarrow \alpha_i) = \frac{P(\alpha_i)}{P(\alpha_{i-1})}$$

$$= \exp\left(\frac{OF_{i-1} - OF_i}{A}\right) \quad \text{for} \quad OF_{i-1} < OF_i \qquad (14a)$$

$$P(\alpha_{i-1} \rightarrow \alpha_i) = 1 \qquad \text{for} \quad OF_{i-1} \geq OF_i. \qquad (14b)$$

Each parameter set is sampled at a random distance and direction from the previously added set, subject to the *a priori* constraints and a specified maximum distance, B. An implementation of the Metropolis algorithm is suggested in Figure 3. Mailhot *et al.* (1997) and Kuczera & Patent (1998) find the Metropolis algorithm to be useful in uncertainty analysis. The Metropolis algorithm can be refined by allowing constants A and B to be updated at intervals, thereby gradually increasing focus on the optima.

Example of calibration

This example aims to demonstrate some of the above approaches to the estimation of a joint *a posteriori* parameter distribution, and establish some relationships and contrasts between them. To make the demonstration manageable, the model is simple and the data are idealised. Attention is drawn to the last paragraph in this
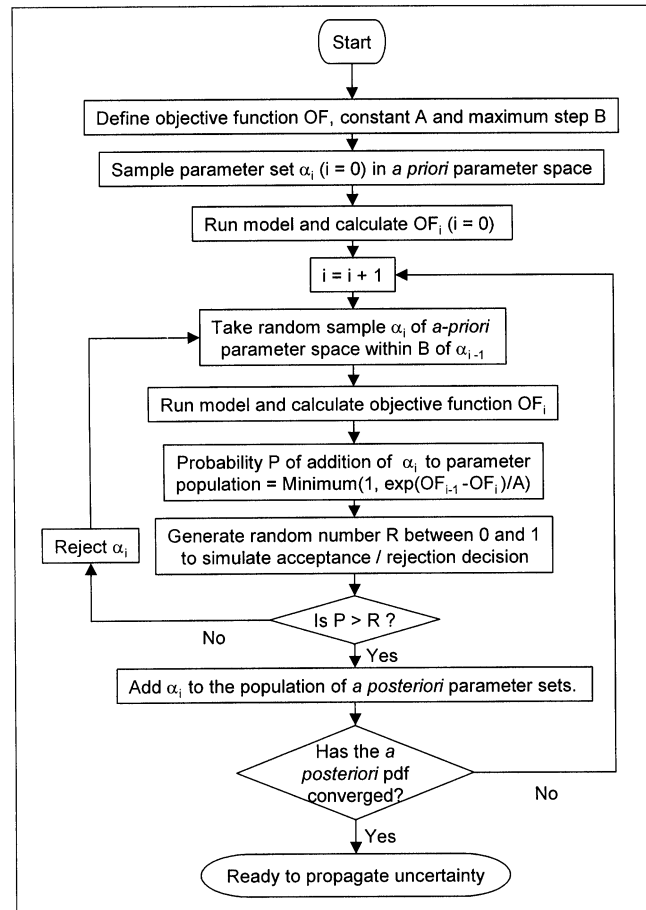

Figure 3 | Metropolis calibration procedure.

section which discusses the example in the context of more complex and practical problems.

A steady state model of organic carbon (BOD) decay and dissolved oxygen (DO) in a river can be described by the Streeter–Phelps equations (Streeter & Phelps 1925):

$$BOD_x = BOD_0 e^{-koc(x/v)} \qquad (15a)$$

$$DO_x = DO_{sat} - \frac{koc \times BOD_0}{kau - koc} [e^{-koc(x/v)} - e^{-kau(x/v)}] -$$
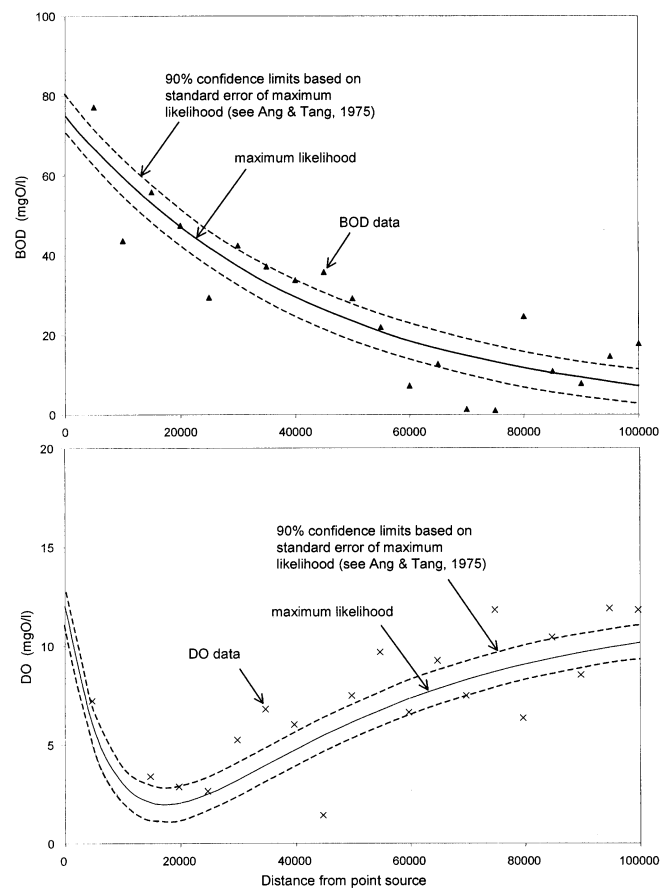
$$(DO_{sat} - DO_0)e^{-kau(x/v)} \qquad (15b)$$

where koc is the BOD decay rate, kau is the oxygen aeration rate, x is the distance downstream from a point pollution source, DO_0 and BOD_0 are the respective concentrations in the river at $x = 0$, v is the average transport

Table 1 | Deterministic parameter values for Streeter–Phelps example

Parameter	Value	Unit
koc	1	d^{-1}
kau	5	d^{-1}
BOD_0	75	mgO/l
DO_0	0	mgO/l
DO_{sat}	12	mgO/l
v	0.5	m/s

Figure 4 | Synthetic data for Streeter–Phelps model.

velocity and DO_{sat} is the concentration of DO at saturation. Synthetic data are generated by the model using the parameter values and boundary conditions in Table 1, and random errors are introduced in $DO(=\varepsilon_{DO})$ from an N(0, 2^2) population, and in BOD ($=\varepsilon_{BOD}$) from an independent N(0, 10^2) population. With 20 data locations spaced at 5 km intervals along a 100 km stretch of river, the synthetic data are illustrated in Figure 4. These data, together with the maximum likelihood model solution, are used to estimate the properties of the distribution of residuals, which are listed (alongside the population properties) in Table 2. This estimation is based on the convergence properties of a sample's maximum likelihood and variance using the Central Limit Theorem (from Ang & Tang 1975). Assuming v and the boundary conditions are known, the parameters to be calibrated are koc and kau.

The synthetically derived error population is randomly sampled and Monte Carlo (MC) calibration of parameters koc and kau proceeds as described in Figure 1. The maximum likelihood set (koc, kau) is found for each data sample using stratified random sampling, using the least sum of squared objective function defined in Equation (5). One thousand samples of the *a priori* parameters are used to estimate the optimum parameter set for each of 200 realisations of the data. That is, 0.2 million model runs are used to derive the posterior parameter distributions. This exercise is repeated with different quantities of synthetic data (i.e. varying the 20 locations shown in Figure 4), with the data error

population distribution kept the same. The comparison of calibrated marginal distributions is shown in Figure 5. Figure 6 gives a similar comparison of the marginal distributions, this time changing the data quality (i.e. varying the expected standard deviation shown in Table 2). Note that the distributions of koc and kau are correlated (correlation coefficient = 0.31), meaning that the *a posteriori* model must be defined by the bivariate distribution of koc and kau as opposed to the marginal distributions shown in Figures 5 and 6. Note also from Figures 5 and 6 that the 'true' values of koc and kau (1 and 5, respectively) do not necessarily correspond to the identified maximum likelihood (see especially the result for 5 data locations in Figure 5). This is because the available data, upon which the MC calibration is founded, are only a *sample* of the true water quality.

Table 2 │ Properties of the DO and BOD data error population

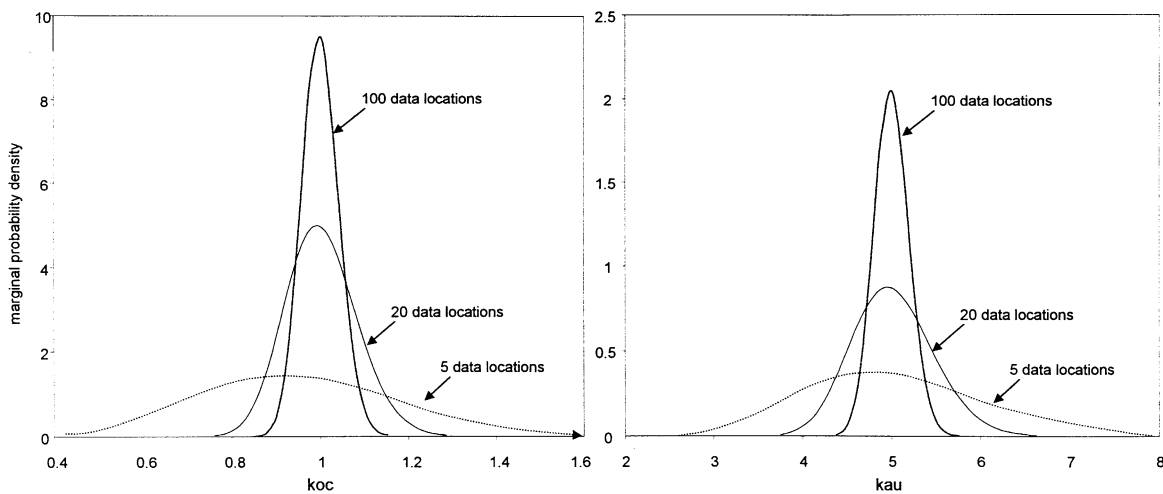|	Property of residuals	Population	Sample	Standard deviation of sampled property
BOD	Mean	0	0	2.14
	Standard deviation	10.00	9.69	5.39
DO	Mean	0	0	0.43
	Standard deviation	2.00	2.17	0.87

Figure 5 │ Calibrated distributions with different sample sizes.

Figure 6 shows that MC gives a significantly uncertain value for the parameter *kau* despite perfect data, which is contrary to intuition. This implies that adequate convergence of the joint parameter distribution has not been achieved using 0.2 million model runs. Whether this is primarily due to the inefficiency of SRS as an optimisation procedure, or due to the limited number of realisations of the data, is not investigated here. However, it is clear that the difficulty of achieving convergence, even for a relatively simple problem such as this, contributes to the approximate nature of the solution.

It is common in environmental monitoring that data are biased descriptors of the true state of the environment. This may be because of heterogeneity which is not recognised in the sampling programme, or because of repeated

laboratory errors, or simply because of physical constraints such as a lower bound of zero. To explore the effect of this, the *DO* data are raised by a random amount between 0 and 5 mgO/l, and to a minimum of zero and the maximum of DO_{sat}. The calibration is done as before, with 20 data locations, and the calibrated parameter distributions are shown in Figure 7. This shows that where significant bias is suspected but unknown, another approach to calibration is required. Note that the parameter uncertainty associated with *kau* is implied to be significantly reduced, contrary to what we would desire. In practice, model structure error is particularly relevant to the Streeter–Phelps model, because it neglects many of the complexities of pollution transport and decay. The effect of model structure error is similar to that of data bias (at
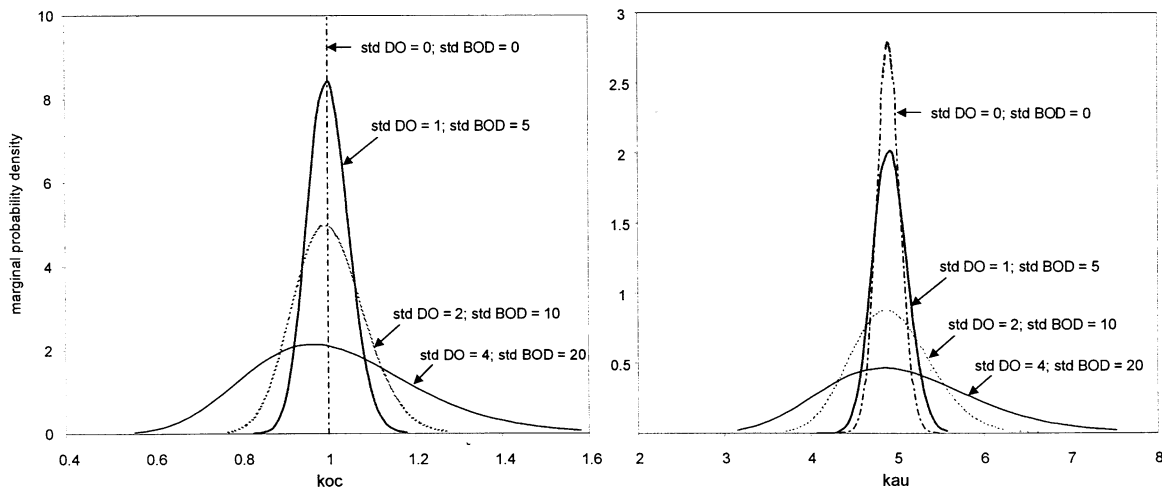
Figure 6 │ Calibrated distributions with different error variances.
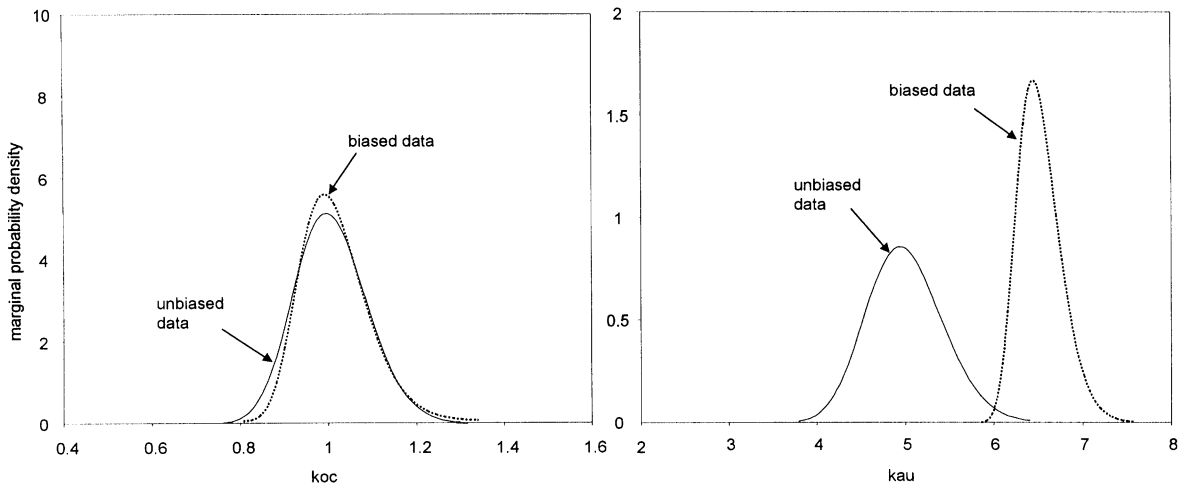

Figure 7 │ Effect of BOD error bias.

least in this case), in that it biases parameter estimates and causes inappropriate reduction in parameter uncertainty.

It has been shown how MC simulation can be used to derive the calibrated marginal parameter distributions by frequency analysis of the sampled maximum likelihood parameters. GLUE offers the opportunity to reduce the computation required by not explicitly accounting for the data sampling error. It is based on point estimates from the joint distribution which are directly applied to uncertainty propagation and, therefore, the non-linear parameter dependencies are implicitly handled. Also, stratified ran-

dom sampling (SRS) gives some control of the location of the point estimates and so coverage of extreme values is assured. Here, GLUE is applied to the previous Streeter–Phelps example using the data sample illustrated by Figure 4. The likelihood function defined in Equation (5) is applied (whereby for this example we are opting not to explore the full generality of GLUE; instead we are maintaining a strictly statistical likelihood estimator), using a total of 2000 random samples of (*koc, kau*). The likelihood equipotentials of the derived point estimates are shown in Figure 8. For comparison with the MC results,
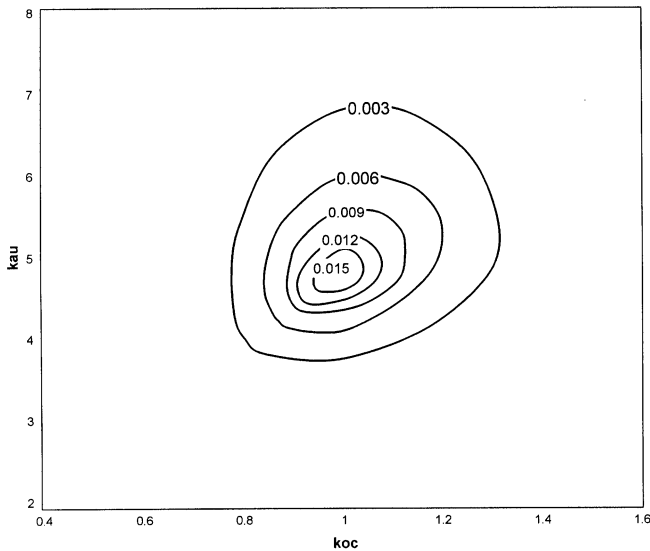
Figure 8 | Equipotentials of point estimate likelihoods using GLUE.

the marginal distributions of *koc* and *kau* are illustrated in Figure 9. Repeated for other data scenarios, the results are summarised in terms of calibrated parameter variances in Figure 10.

The similarity of the GLUE and MC results is striking, considering that the GLUE method does not explicitly account for data sampling error, and has reduced the computation from 0.2 million to 2000 model runs. The theoretical basis for the similarity can be demonstrated at a simple level. Equation (3) is re-expressed as

$$P'_{GLUE} = \frac{1}{K_{GLUE}} \frac{1}{(2\pi(\sigma_m^2+\delta^2))^{0.5N}} \exp[-0.5(N-F)] \quad (16)$$

where P'_{GLUE} is the probability of any parameter set, δ^2 is the variance of the corresponding model result around the maximum likelihood result, σ_m^2 is the error variance around the maximum likelihood result and K_{GLUE} is a standardisation constant. In the MC method, P'_{MC} is the probability of any parameter set, but δ^2 is the variance of the maximum likelihood result for any data realisation around the result for the available data sample. Approximating the standard error of the maxi-

mum likelihood as normally distributed with variance σ_m^2/N (assuming that σ_m is accurate, see Ang & Tang 1975):

$$P'_{MC} = \frac{1}{K_{MC}} \frac{N^{0.5}}{(2\pi\sigma_m^2)^{0.5}} \exp[-0.5\delta^2 N/\sigma_m^2]. \quad (17)$$

As it is known that the integrals of Equations (16) and (17) are both unity, to prove that they give the same result for all parameter samples only requires that the ratio $P'_{GLUE}:P'_{MC}$ is proven to be the same for all δ:

$$\frac{P'_{GLUE}}{P'_{MC}} = \frac{K_{MC}}{K_{GLUE}} \frac{(2\pi\sigma_m^2)^{0.5}}{(2\pi(\sigma_m^2+\delta^2))^{0.5N} N^{0.5}} \times$$
$$\frac{\exp[-0.5(N-F)]}{\exp[-0.5\delta^2 N/\sigma_m^2]}. \quad (18)$$

Amalgamating all terms which are independent of δ into one constant K:

$$\frac{P'_{GLUE}}{P'_{MC}} = K \left[\frac{\exp(\delta^2/\sigma_m^2)}{(\sigma_m^2+\delta^2)} \right]^{0.5N}. \quad (19)$$

Expanding the exponential term into a MacClaurin series, and neglecting terms higher than quadratic, gives,

$$\frac{P'_{GLUE}}{P'_{MC}} = \frac{K \left(\frac{\sigma_m^2+\delta^2}{\sigma_m^2} \right)^{0.5N}}{(\sigma_m^2+\delta^2)^{0.5N}} = \frac{K}{\sigma_m^N} \quad (20)$$

which is constant for all δ. Thus it is shown that Equations (16) and (17) are describing the same probability distribution if δ^4/σ_m^4 and higher-order terms can be neglected. These terms are not negligible if N is very low, but in such cases the assumptions underlying Equations (16) and (17) are not justifiable anyway. Nevertheless, the theory presented here supports the experimental results in Figures 9 and 10, and confirms that GLUE is (usefully) neglecting higher-order uncertainties in this application. Interestingly (but probably not of practical value), Equation (19) provides a basis for exactly reproducing the MC result by adjusting the GLUE result.

The Metropolis algorithm (Figure 3) is observed to further increase the efficiency (in terms of time for
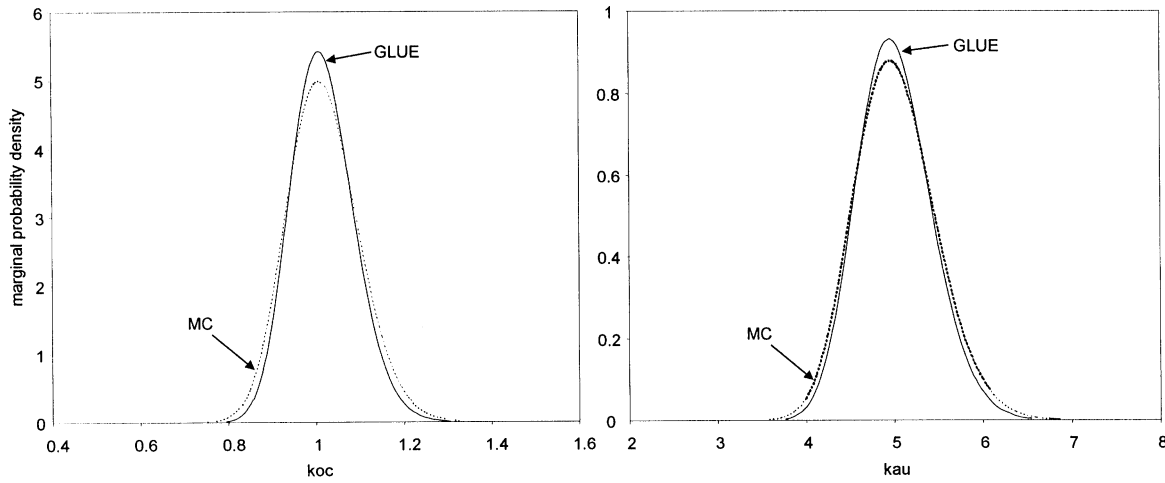
Figure 9 | Comparison of calibrated parameters using GLUE and MC.

convergence of the (*koc, kau*) covariance matrix) of the Streeter–Phelps calibration by up to 60%. The OF is defined as the sum of the variance-weighted squared errors, i.e.

$$OF = \frac{1}{\sigma_{m,DO}^2} \sum_{i=1}^{20} \varepsilon_{i,DO}^2 + \frac{1}{\sigma_{m,BOD}^2} \sum_{i=1}^{20} \varepsilon_{i,BOD}^2 \qquad (21)$$

where $\sigma_{m,DO}$ and $\sigma_{m,BOD}$ are the error population standard deviations (from Table 2), and $\varepsilon_{i,DO}$ and $\varepsilon_{i,BOD}$ are the *i*th residuals of *DO* and *BOD*, respectively. Then, the probability of selecting parameter set α_i pursuant to α_{i-1} is given by Equation (14). *A* is specified as 2, and the maximum permitted step, *B*, is specified as ($B_{koc} = 0.05$, $B_{kau} = 0.25$). The data set illustrated by Figure 4 is used. The converged *koc* and *kau* distributions are almost identical to those obtained using the MC method (Figure 9) and Figure 10 supports this result under a range of data conditions. All the Metropolis results are reproducible with different values of *B*, although the most efficient depends on the convergence criterion. From Equation (13) it is clear that the Metropolis result is sensitive to *A*, and it is not a coincidence that this choice of *A* almost replicates the MC result. Idealising Equation (21) by considering a single response, and using the definitions for Equations (13) and (16):

$$P'_{MET} = \frac{1}{K_{MET}} \exp\left(-\frac{OF_i}{A}\right)$$

$$= \frac{1}{K_{MET}} \exp\left[-\frac{(\delta^2 + \sigma_m{}^2)N}{\sigma_m{}^2 A}\right]$$

$$= \frac{1}{K_{MET}} \exp\left[-\frac{\delta^2 N}{\sigma_m{}^2 A}\right] \exp\left[-\frac{\sigma_m{}^2 N}{\sigma_m{}^2 A}\right]. \qquad (22)$$

Equating this with (17) gives

$$\frac{1}{K_{MET}} \exp\left[-\frac{\delta^2 N}{\sigma_m{}^2 A}\right] \exp\left[-\frac{\sigma_m{}^2 N}{\sigma_m{}^2 A}\right] =$$

$$\frac{1}{K_{MC}} \frac{N^{0.5}}{(2\pi\sigma_m{}^2)^{0.5}} \exp\left[-\frac{0.5\delta^2 N}{\sigma_m{}^2}\right] \qquad (23)$$

and equating the exponents with the δ^2 terms gives $A = 2$. The specification of *A* and the OF used here is generally applicable to an approximation of the standard error of a maximum likelihood model result assuming Gaussian error. As σ_m is not generally known *a priori*, updating of *A* within the algorithm may be useful. Also, *A* and the OF can be designed to give a more useful representation of uncertainty than the standard error, with broader confidence intervals to reflect, for example, unknown model and data bias. Whether this can be done in a robust and meaningful way is a matter for further research. While
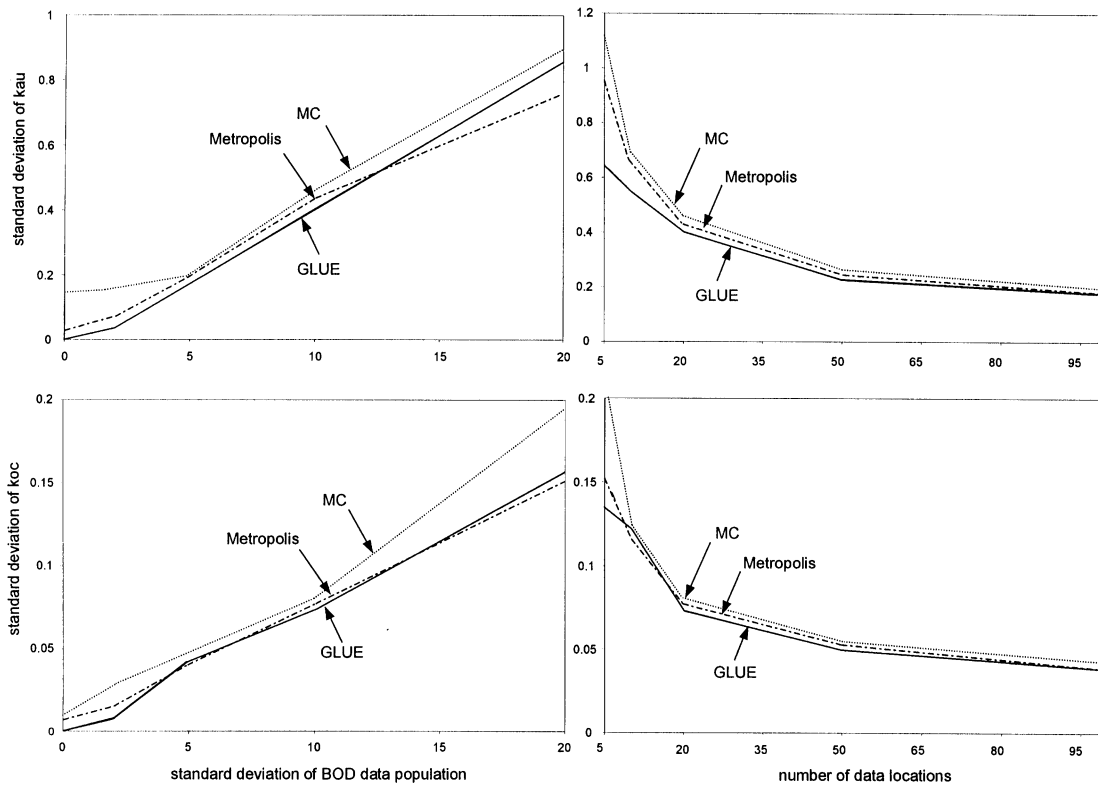
Figure 10 | Comparison of calibrated parameter variances using GLUE, MC and Metropolis.

Metropolis is an adaptive search, and therefore potentially superior to GLUE for finding the maximum likelihood and variance, its sampling of the extreme values is relatively sparse.

Now consider the HSY method of Hornberger & Spear (1980). A set of characteristic system response is defined, with the sampled parameter set given a possibility of 1 ($P(\delta) = 1$), if the model result falls wholly within pre-specified lower and upper limits. For the Streeter–Phelps example, those limits are of DO and BOD (DO_l, DO_u and BOD_l, BOD_u, respectively), i.e.

$$P(\delta) = 1 \text{ for } DO_l < DO < DO_u \cap BOD_l < BOD < BOD_u \quad (24a)$$

$$P(\delta) = 0 \text{ for all other results} \quad (24b)$$

For example, if the upper and lower limits are taken to be the 90% confidence limit defined by the data sample and its maximum likelihood model result, the set of (koc, kau)

defined by Equation (24a) is represented in Figure 11. Note that, as opposed to Figure 8, the set limits defined in Figure 11 are not smooth due to the discontinuous nature of Equation (24). The HSY method is potentially more robust to model error and data bias than statistically based likelihood methods because unreliable results, such as those illustrated in Figure 7, can be avoided with appropriate specification of the upper and lower bounds of characteristic response. Of course, improvement in reliability is at the expense of a less specific description of uncertainty.

Notwithstanding its demonstrative limitations (see below), this example compares classical Monte Carlo simulation, GLUE and Metropolis and shows that these methods are not fundamentally different insofar as they can produce the same calibration results, given consistent objective functions. The question of which is more efficient depends upon the case. Monte Carlo sampling of the data is a rigorous method of sampling from a known error distribution, and can be modified to a resampling
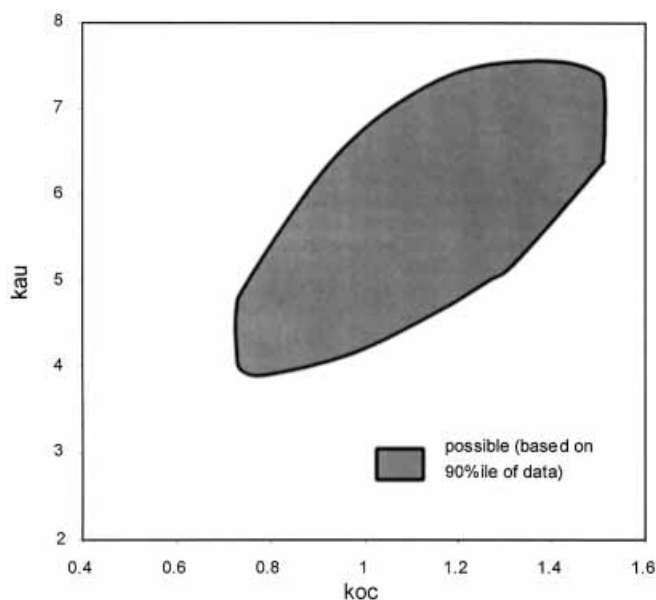
Figure 11 │ The (*koc, kau*) uniform possibility distribution.

illustrated by Figure 8, is well behaved. Many practical problems involve multi-modal responses together with discontinuities derived from the discontinuities in the model structure, again increasing the difficulty of convergence. Thirdly, Equation (15) is an analytical solution to the Streeter–Phelps model, which is solved easily and quickly, which facilitates Monte Carlo methods. Models of the environment are more often in the form of systems of differential equations for which approximate numerical solutions are required and computational demands are relatively high. While computer power is continuously increasing and parallel processing facilities are available, computation time remains a limitation in model calibration and uncertainty analysis. Lastly, the data have been synthesised from a normal population of residuals which are uncorrelated and have zero mean. Only a nominal look at the effects of data bias has been included.

scheme if the error distribution is unknown. GLUE avoids sampling of data and, if used with statistical likelihood estimators, can be an extremely good approximation to the more rigorous data sampling schemes. In addition, it has more general application than used in the example (see below). Metropolis has the advantage of adapting the response surface, thereby generally locating the parameter set mode(s) to within some tolerance in a shorter time. However, this is at the expense of the reliability of extreme results. All these approaches to calibration require careful design of the OF, as this defines the variance of the *a posteriori* joint parameter distribution. The HSY algorithm was briefly explored, and it was shown how it might be used to cover for the effects of model and data biases, at the cost of a less discriminating description of uncertainty.

With regard to more complex and more realistic environmental modelling, the above example has several important limitations. Firstly, it only has two interdependent parameters, while many models have significantly more. In such cases converging the *a posteriori* joint distribution would be expected to be much more difficult, perhaps requiring many thousands of model runs (e.g. Thyer *et al.* 1999), depending on the strength and nature of the interactions. Secondly, the response surface, which is

UNCERTAINTY PROPAGATION

Uncertainty propagation in this context means propagating the calibrated parameter joint distribution to a stochastic result. Methods of propagating probability distributions can be classified as sampling methods, variable transformation methods, point estimation methods and variance propagation methods. An alternative to probability theory is the theory of possibility (Zadeh 1978). Each of these approaches is discussed here. In this discussion, the general term 'random variables' includes uncertain parameters, boundary conditions and stochastic results.

Monte Carlo methods

Monte Carlo (MC) simulation applied to uncertainty propagation means generating discrete parameter sets according to their probability or possibility distribution, and running a simulation using each set. Alternatively, the population or point estimates which were derived during calibration can be recalled, thereby avoiding the need for assumptions regarding the form of the distribution. The

results of multiple simulations give a close approximation to the analytical form of the probability density function (PDF) using frequency analysis, and any model can be easily included in such a framework with minimal input from the modeller. For these reasons, MC is a well used method of uncertainty propagation. The main disadvantage of MC is that a great number of model runs may be required to reliably represent all probable results, especially when there is a number of random variables. Methods of estimating a preferred number of samples are available (e.g. Cochran 1977), although this also depends on the convergence or divergence during propagation and therefore is case specific (Tellinghuisen 2000). Variance reduction techniques, such as stratified random sampling (SRS) and Latin hypercube sampling (see MacKay *et al.* 1979), are often used to improve efficiency.

First-order and point estimate approximations

First-order variance propagation is the most common method of uncertainty propagation (Beck 1987). If a function $Z = f(X)$, $X = (x_1, x_2, \ldots x_F)$ is approximated by a first-order Taylor series expansion around the expected X, μ_x, then

$$\mu_Z = f(\mu_X) \tag{25a}$$

$$\sigma_Z^2 = \Delta(Z)^T \Psi(X) \Delta(Z) \tag{25b}$$

where $\Delta(Z)$ is the $F \times 1$ matrix of derivatives of Z with respect to X; $\psi(X)$ is the $F \times F$ covariance matrix of $(x_1, x_2, \ldots x_F)$ and μ_Z and σ_Z^2 are the mean and variance of Z. This is a linear approximation of uncertainty propagation which is only completely reliable for linear models. The accuracy of this method for non-linear models can be improved by using a higher-order Taylor series expansion, but this becomes computationally demanding, especially if the derivative values are calculated numerically. Variance propagation is a useful method for models which can approximated by quasi-linearisation (see Gelb 1974), i.e. a series of localised linear functions. For example, Kitanidis & Bras (1980) applied variance propagation to a quasi-linearised hydrologic model.

Rosenblueth's point estimation method for symmetric and non-symmetric variable distributions (Rosenblueth 1981) aims to reduce the computational demands of variance propagation by eliminating the calculation of derivatives. The PDF of each random variable is represented by p discrete points, located according to the first, second and third moments of the PDF. The joint PDF of F random variables is represented by the array of projected points. Therefore, p^F points are used. Each point is assigned a mass according to the third moment and the correlation matrix. All points are propagated discretely to p^F solutions and the first moment is the weighted average; the second moment, that of the squares; and the third moment, that of the cubes. Most usually a two-point scheme is used whereby 2^F points are required. For symmetrical distributions for $F > 2$, the number of evaluations can be reduced from 2^F to 2^F by using Harr's point estimation method (Harr 1989). Harr's method is a useful improvement on Rosenblueth's, but is limited by the necessity of symmetrical distributions. A similar approach which allows for skewed distributions but not correlations is described by Hong (1998). Protopapas & Bras (1990) have applied Rosenblueth's two-point method to a rainfall-runoff model and Yeh *et al.* (1997) have similarly applied Harr's method, and a useful review of all these point estimate methods is given by Christian & Baecher (1999).

It is worth noting that some Monte Carlo-based methods of calibration, for example GLUE (Beven & Binley 1992), are randomised point estimate methods. In GLUE, many random samples of the *a priori* parameter space are assigned GLUE likelihoods, which then become point estimates for the uncertainty propagation stage. Unlike Rosenblueth's method, it is generally assumed that there are enough points to derive the model output PDF, not just the lower moments.

Possibility theory

Possibility theory (Zadeh 1978) offers a robust alternative to propagation of probability distributions. To illustrate this, let $f(X, Y)$ be a function which is strictly increasing or decreasing with respect to both variables X and Y, and let X and Y have possibility distributions which rise internally

to a single peak or plateau. Then, using the rules of possibility in Equations (12a) and (12b), the two values of $X(=x_{lp}, x_{up})$ and $Y(=y_{lp}, y_{up})$ with possibility P (commonly called the P-level α-cut of X and Y) define the two values of $f(X,Y)$ with possibility P. For example, if $\partial f/\partial X$ is positive for all X and $\partial f/\partial Y$ is negative for all Y, then

$$f(X,Y)_{up} = f(x_{up}, y_{lp}) \qquad (26a)$$

$$f(X,Y)_{lp} = f(x_{lp}, y_{up}) \qquad (26b)$$

This method can be extended to problems with many uncertain parameters so long as the aforementioned 'increasing–decreasing' and 'single-peak' conditions are met for each. While an infinite number of α-cuts are required for the exact solution to non-linear propagation (Wierman 1996), an approximation of the propagated possibility distribution can be made with a small number of computations. The associated difficulties and limitations should be recognised. Firstly, special attention must be given to the method of calibration in order to derive meaningful parameter possibility distributions. Secondly, the possibility is greater than the probability at all points (Zadeh 1978), and so the former is a less specific descriptor of uncertainty. Thirdly, if parameter α-cuts are to be used, prior knowledge of the sensitivity of the results to the parameters is required. Lastly, there remains the problem of parameter interdependence which, as in probability theory, complicates the analysis, generally requiring that the P-level α-cut be defined by a large sample of parameter sets.

Propagation of the Streeter–Phelps model parameters

The joint (*koc, kau*) distribution previously identified using GLUE (Figure 8) is propagated to give spatially varying distributions of BOD and DO. Again, the boundary conditions defined in Table 1 are used. Firstly, each of the 2000 point estimates of (*koc, kau*) is propagated through the Streeter–Phelps model, then the first-order variance and Rosenblueth two-point methods are applied, using the covariance matrix of (*koc, kau*) which is derived from the GLUE point estimates. It is observed that the
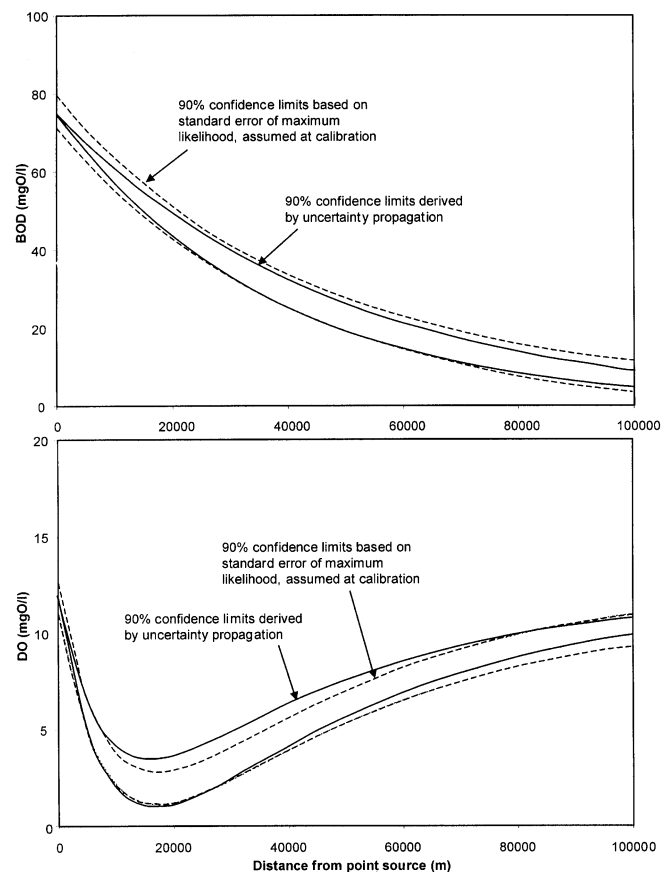

Figure 12 | Propagated uncertainty for Streeter–Phelps model.

three alternative methods give (practically) identical results for the first three moments. The numerical efficiency of the first-order variance and Rosenblueth two-point methods is proven, despite the apparent non-linearity of the model with respect to *koc* and *kau*. In fact, the model is only significantly non-linear at low values of DO, and the performance of the first-order method deteriorates with either increased BOD loading or increased data uncertainty. Propagation of the possibility distribution illustrated in Figure 11 replicates the 90% confidence limits derived using the other methods (assuming normality), although this result is specific to the definitions used in the derivation of Figure 11 (see Equations (24a,b)). None of the methods can reproduce the 90% confidence limits on the data error upon which the calibration was founded, illustrated in Figure 12.

Results are constrained by the boundary conditions, irrespective of parameter uncertainty. Therefore, if they are not precise, the modeller must treat the boundary conditions as random variables as well as the parameters. Furthermore, it may be observed from Figures 4 and 12 that the 90% confidence limits contain much less than 90% of the data. Using the multiplicative likelihood estimator of Equation (5) has meant the limits represent the *uncertainty in the maximum likelihood solution*, not the variance of the data.

The example of the Streeter–Phelps model has illustrated that alternative methods of estimation and propagation of parametric uncertainty can lead to practically the same result. However, the example is too simple to fully show the limitations of the reviewed methods. At the calibration stage, unknown model structure and data bias may mean that statistical interpretation of the model residuals is not useful, and a more robust approach must be sought, by subjective evaluation. At the propagation stage there may be strong non-linear dependency of parameters, which must be approximated by a covariance or correlation coefficient in Rosenblueth's method and the first-order variance method, leading to a poor approximation of prediction uncertainty. Practical environmental models often include cyclic, non-continuous or otherwise highly non-linear processes which will test all methods much more severely than was attempted here. Practical models may take considerable processing time and have many uncertain parameters. Achieving reliable results using MC simulation may be computationally expensive, although this is of less concern where parallel processing facilities are available.

CONCLUSIONS

Review summary

Imprecision in environmental modelling stems from the inevitable approximate nature of the models, and from the inevitable difficulty of identifying a single 'best' model, given the limitations in our prior knowledge and in the information retrievable from field data. In general, it may be said that the natural environment is too complex, with too many heterogeneities and apparently random influences, to be usefully described without including some estimation of uncertainty. The inclusion of uncertainty analysis adds to a conventional modelling exercise in two main ways. Firstly, the calibration of model parameters involves identification of parameter distributions rather than single parameter values. Secondly, the parameter distributions (and alternative model structures if used) are propagated to stochastic rather than deterministic results.

A variety of applications of Monte Carlo (MC) simulation can provide detailed information about the calibrated parameter distributions. A traditional application of MC simulation is to optimise the parameters with respect to each of a multitude of realisations of the data, thereby giving a sample of optima. If data are of limited quality, this sample converges slowly, and reliable results may not be achieved within an acceptable time. Generalised Likelihood Uncertainty Estimation (GLUE) requires the modeller to design a likelihood measure which is used in the calibration. This design may be based on classical likelihood estimators, in which case GLUE is seen to significantly improve the efficiency of traditional MC simulation while giving similar results. Alternatively, the GLUE likelihood measure may include subjective interpretation of the data as well as the potential for model structure error, in order to give more robust results. GLUE is founded on the principles of Bayesian estimation, and in this regard is especially useful for updating model uncertainty estimation as new data, knowledge and models come to light. The Metropolis algorithm is a Monte Carlo Markov chain procedure which has proven useful in parameter uncertainty estimation. In some cases it is a more efficient method of deriving parameter distributions than GLUE, although it cannot give such detailed information about extreme values. The theoretical and practical similarity of the traditional MC, GLUE and Metropolis methods of calibration, under idealised conditions, has been demonstrated.

The computational burden of MC methods has, in the past, been a major limitation. However, MC simulation is ideally suited to parallel processing. Also, a number of variance reduction techniques are available to improve efficiency. Stratified random sampling is particularly

valuable in environmental modelling because it guarantees a number of extreme value samples, and that all significant parameter correlations are efficiently identified and propagated. Latin hypercube sampling further improves the representation of individual parameters at the expense of a representation of their interactions.

For propagation of parametric uncertainty, Monte Carlo methods are generally the most useful. Various other methods give relatively efficient estimation of the lower moments of propagated probability distributions (e.g. Rosenblueth's two-point method, first-order variance propagation). However, the applicability of these methods generally decreases as model complexity grows, and the results should be subject to verification.

A look to the future

In a society which is increasingly reliant upon and confident with computer technology, and which is increasingly sympathetic to the need for environmental protection, the use of computer models of the environment is bound to flourish. In the future, the stakes involved with environmental management are likely to increase and (as in other professions where the stakes are perceived as high) risk–benefit management will become a key concept. The risks associated with environmental interventions mainly stem from a lack of knowledge about environmental responses at regional to global scales. Using stochastic modelling methods, we are increasingly able to include this lack of knowledge as an essential input to environmental risk management. Continuing advances in computational resources, for example parallel processing technology and powerful desktop computers, will allow more widespread and economical application of the Monte Carlo-based methods described in this paper. Also, there is scope for the development and application to environmental modelling of more advanced algorithms for the identification of model uncertainty (for example, genetic algorithms, shuffled complex evolution and Markov chain methods).

It was argued in this paper that complex models tend to be less identifiable because it is more difficult to identify which model components are causing the observed responses. Therefore, uncertainty and modelling expense can be reduced by using parsimonious models which aim to include only the principal components affecting the system under observed conditions. On the other hand, the predictive power of a model lies largely in its ability to explore changes in the principal components, for which hypotheses of future behavioural changes and suitable model structures are needed. Clearly, it is important that modellers have the expertise and information needed to select an appropriate model structure. This may be achieved by development and dissemination of modelling toolkits, by which modellers can take an experimental approach to model structure identification (e.g. Wagener *et al.* 2001). Such tools also provide a way forward for incorporating alternative structural hypotheses into the uncertainty estimation.

If there is inadequate information with which to identify a model, the uncertainty in the model is defined by *a priori* research, which would generally include a review of the parameter distributions identified for previous comparable modelling studies. It can be argued, then, that the modelling profession would benefit from formal compendiums of modelling studies. Further to this, more research is required into the problem of model regionalisation—relating model structures and *a priori* parameter distributions to the characteristics of the systems under study. For example, clusters of models (structures and corresponding parameter distributions) can be recommended for a problem with certain attributes. Although regionalisation approaches are established in some modelling disciplines (e.g. rainfall-runoff modelling), the gathering and processing of world-wide experience in identifying models to system characteristics is, generally speaking, a task for the future decades.

From this review, it is clear that objective function (OF) design is difficult because traditional likelihood estimators are not robust to model structure and data bias, while the significance of more subjective measures of model performance is not easy to define. Two on-going approaches to these problems seem likely to become prominent in the future. Firstly, multi-objective optimisation (whereby two or more OFs are applied to the same calibration problem) can be used to explore the robustness of the optimum model to OF design, and to explore the

inability of the model to achieve multiple objectives simultaneously (e.g. Gupta *et al.* 1998). As the latter is the symptom of model structural error, this provides a basis for model structure identification (e.g. Wagener *et al.* 2002). Multi-objective optimisation was excluded from this review because successful applications have been to problems which are well defined by observations. It is a matter for future research how such methods can be employed to improve robustness of uncertainty analysis of less well defined systems.

A look to the future of uncertainty analysis in environmental modelling is not complete without a discussion of the crucial role of environmental monitoring. As previously stated, a model is a collection of hypotheses of system behaviour conditioned by observations of the system, which usually come from formal monitoring programmes. In previous decades most monitoring has usually been fixed in frequency and location for regulation purposes, rather than designed to encapture the system responses which are required for model identification. With increasing regulatory motivation, more resources are likely to be made available for model-oriented monitoring. This may include *in situ* monitors at strategic locations to give representative and intensive temporal measurements, and continued application of satellite imagery for spatial representations (e.g. Franks & Beven 1999). Access to new data should be encouraged, for example along the lines recently implemented by the USGS (Benson & Faundeen 2000).

ACKNOWLEDGEMENTS

NOMENCLATURE

M	notional model result including error
O	notional measurement including error
N	number of residuals

F	number of model parameters
R	number of model responses
S	number of samples of parameter sets
D	number of samples of data sets
K	constant as defined in text
$P(\ )$	probability or possibility as defined in text
X, Y	arbitrary independent random variables
A, B	constants used in Metropolis algorithm
Z	arbitrary variable Z dependent on X
$\Delta(Z)$	matrix of derivatives of Z with respect to X
$\Delta(Z)^{\mathrm{T}}$	transpose of $\Delta(Z)$
$\psi(X)$	covariance matrix of X
μ_X	expected value of X
x	distance downstream in Streeter-Phelps model
v	average water velocity in Streeter-Phelps model
BOD	biochemical oxygen demand concentration
DO	dissolved oxygen concentration
koc	oxidation rate of BOD
kau	reaeration rate of DO
ε	model residual
$\varepsilon_1, \varepsilon_2, \varepsilon_3$	a series of model residuals, or components of residuals
σ_m^2	variance of residuals around optimum model result
δ^2	variance of model solution around optimum model result
α	a parameter vector
α'	an optimum parameter vector

ABBREVIATIONS

OF	objective function
PDF	probability density function
SRS	stratified random sampling
ARS	adaptive random search
GLUE	Generalised Likelihood Uncertainty Estimation
HSY	Hornberger–Spear–Young algorithm

REFERENCES

Ang, A. H. S. & Tang, W. H. 1975 *Probability Concepts in Engineering Planning and Design, Vol.1- Basic Principles.* John Wiley, New York, pp. 56, 222, 232–248.

Ang, A. H. S. & Tang, W. H. 1984 *Probability Concepts in Engineering Planning and Design, Vol.2—Decision, Risk and Reliability*. John Wiley, New York, p. 274.

Beck, M. B. 1987 Uncertainty in water quality models. *Wat. Res. Res.* **23**(8), 1393.

Benson, M. G. & Faundeen, J. L. 2000 USGS remote sensing and geoscience data: using standards to serve us all. *Proceedings of International Geoscience and Remote Sensing Symposium (IGARSS)*. The Institute of Electrical and Electronics Engineers (IEEE), Hawaii, vol 3, pp. 1202–1204.

Berthouex, P. M. & Brown, L. C. 1994 *Statistics for Environmental Engineers*. CRC Press, Boca Raton, FL, p. 210.

Beven, K. J. & Binley, A. M. 1992 The future of distributed models; model calibration and predictive uncertainty. *Hydrol. Processes* **6**, 279–298.

Bierman, V. J. & Dolan, D. M. 1986 Modeling of phytoplankton in Saginaw Bay: II. Post-audit phase. *ASCE J. Environ. Engng.* **112**(2), 415–429.

Binley, A. M., Beven, K. J., Calver, A. & Watts, L. G. 1991 Changing responses in hydrology: assessing the uncertainty in physically based model predictions. *Wat. Res. Res.* **27**, 1253–1261.

Box, G. E. P. & Jenkins, G. E. 1970 *Time Series Analysis Forecasting and Control*. Holden-Day, San Francisco, p. 208.

Chen, J. & Wheater, H. 1999 Identification and uncertainty analysis of soil water retention models using lysimeter data. *Wat. Res. Res.* **35**(8), 2401–2414.

Christian, J. T. & Baecher, G. B. 1999 Point-estimate method as numerical quadrature. *J. Geotech. Geoenviron. Engng.* **125**(9), 779–786.

Cochran, W. G. 1977 *Sampling Techniques*, 3rd edn. John Wiley, New York, p. 72.

Duan, Q., Gupta, V. K. & Sorooshian, S. 1993 A shuffled complex evolution approach for effective and efficient global minimization. *J. Optim. Theory Appl.* **76**(3), 501–521.

Fonseca, C. M. & Fleming, P. J. 1995 An overview of evolutionary algorithms in multi-objective optimization. *Evolut. Comput.* **3**(1), 1–16.

Franks, S. W. & Beven, K. J. 1999 Conditioning a multiple-patch SVAT model using uncertain time-space estimates of latent heat fluxes as inferred from remotely sensed data. *Wat. Res. Res.* **35**(9), 2751–2761.

Freer, J., Beven, K. & Ambroise, B. 1996 Bayesian estimation of uncertainty in runoff prediction and the value of data: an application of the GLUE approach. *Wat. Res. Res.* **32**(7), 2161.

Gelb, A. 1974 *Applied Optimal Estimation*. MIT Press, Cambridge, MA.

Gupta, V. K., Sorooshian, S. & Yapo, P. O. 1998 Toward improved calibration of hydrologic models: multiple and noncommensurable measures of information. *Wat. Res. Res.* **34**(4), 751–763.

Harr, M. E. 1989 Probabilistic estimates for multi-variate analyses. *Appl. Math. Model.* **13**(5), 313–318.

Holland, J. H. 1975 *Adaption in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor, MI.

Hong, H. P. 1998 An efficient point estimate method for probabilistic analysis. *Reliab. Engng. Syst. Safety* **59**(3), 261–267.

Hornberger, G. M. & Spear, R. C. 1980 Eutrophication in Peel Inlet, 1. Problem-defining behaviour and a mathematical model for the phosphorus scenario. *Wat. Res.* **14**, 29–42.

Kitanidis, P. K. & Bras, R. L. 1980 Real-time forecasting with a conceptual hydrologic model: 1. Analysis of uncertainty. *Wat. Res. Res.* **16**(6), 1025–1033.

Kuczera, G. & Parent, E. 1998 Monte Carlo assessment of parameter uncertainty in conceptual catchment models. *J. Hydrol.* **211**(1), 69–85.

MacKay, M. D., Beckman, R. J. & Conover, W. J. 1979 A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics.* **21**(2), 239–245.

Mailhot, A., Gaume, E. & Villeneuve, J. P. 1997 Uncertainty analysis of calibrated parameter values of an urban storm water quality model using Metropolis Monte Carlo algorithm. *Wat. Sci. Technol.* **36**(5), 141–148.

Melching, C. 1995 Reliability estimation. In: *Computer Models of Watershed Hydrology*, (ed. Singh, V. P.). Water Resource Publications, New York pp. 69–118.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, F. 1953 Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21**, 1087.

Mulligan, A. E. & Brown, L. C. 1998 Genetic algorithms for calibrating water quality models. *ASCE J. Environ. Engng.* **124**(3), 202–211.

Protopapas, A. L. & Bras, R. L. 1990 Uncertainty propagation with numerical models for flow and solute transport in the unsaturated zone. *Wat. Res. Res.* **26**(10), 2463–2474.

Romanowicz, R., Beven, K. & Tawn, A. 1994 Evaluation of predictive uncertainty in non-linear hydrological models using a Bayesian approach. In: *Statistics For The Environment 2*, (eds. Barnett, V. & Turkman, K. F.). John Wiley, Chichester, pp. 297–318.

Rosenblueth, E. 1981 Two-point estimates in probabilities. *Appl. Math. Model.* **5**, 329–335.

Rubinstein, R. Y. 1981 *Simulation and the Monte Carlo Method*. John Wiley, New York, p. 240.

Rutenbar, R. A. 1989 Simulated annealing algorithms: an overview. *IEEE Circuits Dev. Mag.* Jan, 19–26.

Shao, J. 1996 Bootstrap model selection. *J. Am. Statist. Assoc.* **91**(434), 655–665.

Shao, J. & Tu, D. 1995 *The Jackknife and Bootstrap*. Springer-Verlag, Berlin, p 2.

Sorooshian, S. & Dracup, J.A. 1980 Stochastic parameter estimation procedures for hydrologic rainfall-runoff models: correlated and heteroscedastic error cases. *Wat. Res. Res.* **16**(2), 430–442.

Sorooshian, S. & Gupta, V. K. 1995 Model calibration. In: *Computer Models of Watershed Hydrology*, (ed. Singh, V. P.). Water Resources Publications, New York pp. 23–68.

Spear, R. C. & Hornberger, G. M. 1980. Eutrophication in Peel Inlet 2. Identification of critical uncertainties via generalized sensitivity analysis. *Wat. Res.* **14**(1), 43–49.

Streeter, H. W. & Phelps, E. B. 1925. A study of the pollution and natural purification of the Ohio River, III. Factors concerned

in the phenomena of oxidation and reareation. *US Public Health Service, Public Health Bulletin*, **146**.

Tellinghuisen, J. A. 2000 Monte Carlo study of precision, bias, inconsistency, and non-Gaussian distributions in non-linear least squares. *J. Phys. Chem*. **104**(12), 2834–2844.

Thomann, R. V. 1998. The future 'golden age' of predictive models for surface water quality and ecosystem management. *ASCE J. Environ. Engng* **124**(2), 94–103.

Thyer, M., Kuczera, G., Bates, B. C. & Bryson C. 1999 Probabilistic optimization for conceptual rainfall-runoff models: a comparison of the shuffled complex evolution and simulated annealing algorithms. *Wat. Res. Res.* **35**(3), 767–773.

Tung, Y. K. 1996 Uncertainty and reliability analysis. In: *Water Resources Handbook*, (ed. Mays, L. W.). McGraw-Hill, New York, pp. 7.1–7.65.

Valdes, J. B., Rodriguez-Huras, I. & Vicens, G. J. 1980 Choosing among hydrologic regression models 2. Extensions to the standard model. *Wat. Res. Res*. **16**(3), 507–516.

Van Straten, G. & Keesman, K. J. 1991 Uncertainty propagation and speculation on projective forecasts of environmental change: a lake eutrophication example. *J. Forecasting* **10**, 163–190.

Wagener T., Boyle D. P., Lees M. J., Wheater H. S., Gupta H. V. & Sorooshian S. 2001 A framework for the development and application of hydrological models. *Hydrol. Processes* in press.

Wagener, T., McIntyre, N., Lees, M. J., Wheater, H. S. and Gupta, H. V. 2002. Reducing uncertainty in conceptual rainfall-runoff modelling: dynamic identifiability analysis. *Hydrol. Earth Syst. Sci.* **5**(1), 13–26.

Weglarczyk, S. 1998 Interdependence and applicability of some statistical quality measures for hydrological models. *J. Hydrol*. **206**(1–2), 98–103.

Wheater, H. S., Bishop, K. H. & Beck, M. B. 1986 The identification of conceptual hydrological models for surface water acidification. *Hydrol. Processes* **1**, 89–109.

Whitehead, P. G. & Hornberger, G. M. 1984 Modelling algal behaviour in the River Thames. *Wat. Res*. **18**(18), 945–953.

Wierman, M. J. 1996 Assessing fuzzy sets and the potential of possibility theory. *Inf. Sci*. **88**(1–4), 247–261.

Yeh, K., Yang, J. C. & Tung, Y. K. 1997 Regionalization of unit hydrograph parameters: 2. Uncertainty analysis. *Stochast. Hydrol. Hydraul*. **11**, 173–192.

Zadeh, L.A. 1978 Fuzzy sets as a basis for the theory of possibility. *Fuzzy Sets Syst*. **1**, 3–28.