

Uncertainty and sensitivity analysis techniques for hydrologic modeling

Srikanta Mishra

ABSTRACT

Formal uncertainty and sensitivity analysis techniques enable hydrologic modelers to quantify the range of likely outcomes, likelihood of each outcome and an assessment of key contributors to output uncertainty. Such information is an improvement over standard deterministic point estimates for making engineering decisions under uncertainty. This paper provides an overview of various uncertainty analysis techniques that permit mapping model input uncertainty into uncertainty in model predictions. These include Monte Carlo simulation, first-order second-moment analysis, point estimate method, logic tree analysis and first-order reliability method. Also presented is an overview of sensitivity analysis techniques that permit identification of those parameters that control the uncertainty in model predictions. These include stepwise regression, mutual information (entropy) analysis and classification tree analysis. Two case studies are presented to demonstrate the practical applicability of these techniques. The paper also discusses a systematic framework for carrying out uncertainty and sensitivity analyses.

Key words | hydrologic, modeling, sensitivity, techniques, uncertainty

Srikanta Mishra
INTERA Inc.,
1812 Centre Creek Drive, Suite 300,
Austin TX 78754,
USA
Tel.: +1 512 425 2071
E-mail: smishra@intera.com

BACKGROUND AND SCOPE

Engineers and scientists dealing with the modeling and management of hydrologic systems are routinely confronted with uncertainty caused by incomplete knowledge and/or natural variability. Regardless of the source, uncertainties impact all facets of hydrologic data collection and modeling, i.e. (a) conceptual and numerical model development, (b) estimation of model parameters and (c) quantification of historical and/or future model stresses.

Traditional deterministic analysis of uncertainty in hydrologic models often involves the use of best-guess or worst-case assumptions about model inputs to quantify their impacts on model predictions. Alternatively, a set of optimistic and pessimistic values is sometimes utilized to provide upside and downside forecasts around a reference scenario. Typically, this simplistic approach is not capable of dealing with complex problems where the system response is far from linear or where significant correlations exist between model parameters. Systematic combinations

of optimistic and pessimistic values may also lead to confidence intervals that are too wide (resulting in over-design), and whose reliability is difficult to assess.

Recently, there has been greater interest in the use of probabilistic techniques for formal uncertainty assessment in hydrologic modeling. Such a change in mind set has been spurred to some extent by the recognition that making better decisions under uncertainty requires some understanding of the range and likelihood of possible outcomes—as opposed to basing decisions on traditional single-point estimates.

Before diving into the nitty-gritty of modeling under uncertainty, it is useful to outline the key elements of a systematic framework for dealing with uncertainty. These key elements are:

- *uncertainty characterization*—which involves fitting and/or assigning marginal and joint distributions to uncertain model inputs,

- *uncertainty propagation*—which involves translating the uncertainty in model inputs into the corresponding uncertainty in model outputs, and
- *uncertainty importance assessment*—which involves determining the key drivers of output uncertainty.

In the context of probabilistic modeling, “uncertainty analysis” generally includes both uncertainty characterization and propagation, whereas “sensitivity analysis” generally refers to the assessment of uncertainty importance. Uncertainty analysis thus allows capturing the full spectrum of information regarding uncertain and variable factors, and estimating distributions around model predictions. Sensitivity analysis, on the other hand, enables identifying key input parameters that contribute the most to the model’s predictive uncertainty.

The objective of this paper is to provide an overview of practical uncertainty and sensitivity analysis techniques that can be readily applied to hydrologic modeling studies. A brief theoretical discussion is provided for each of the techniques, along with illustrative examples. Finally, two case studies are presented to demonstrate the challenges in applying these methods to real-life problems.

UNCERTAINTY ANALYSIS TECHNIQUES

As noted earlier, uncertainty analysis involves characterizing uncertainty in model inputs via probability distributions, and propagating these via system models into corresponding uncertainties in model outcomes. In hydrologic applications, uncertainty analysis is often taken to be synonymous with Monte Carlo simulation (MCS). In the MCS methodology, multiple model calculations are performed with parameter values drawn randomly from specified probability distributions to compute the probability distribution of model outputs of interest (e.g. Ang & Tang 1984). Although MCS provides the greatest versatility in uncertainty propagation studies, it may not be the most efficient when: (a) parameter uncertainty is poorly defined, (b) forward models are computation intensive or (c) outcomes of interest are limited in number (Mishra 2003).

In this section, a systematic approach is first outlined for MCS applications. The motivation here is to reinforce the point that there is more to MCS than simple random

sampling, multiple model runs and aggregation of results. Similar ideas have been recently presented in the risk assessment literature (EPA 1997). Next, alternative uncertainty analysis techniques that complement MCS are described—along with their advantages and limitations.

Monte Carlo simulation (MCS)

Application of the MCS methodology for uncertainty analysis involves the following steps:

- Selecting imprecisely known model input parameters to be sampled.
- Assigning ranges and probability distributions for each of these parameters.
- Generating many sample sets (realizations) with random values of model parameters.
- Running the model for all realizations to estimate uncertainty in model outcomes.

The goal of the first step, i.e. selection of uncertain inputs, is to identify and retain only those input variables that have the greatest impact on the outcomes of interest. It should be carried out using subjective judgment, standard one-parameter-at-a-time sensitivity analysis or randomized one-parameter-at-a-time screening (e.g. Morris 1991). Eliminating redundant uncertain inputs from the sampled set generally helps focus data-collection efforts and improves the stability and reliability of probabilistic model results (EPA 1997). It also facilitates robust statistical model building of input–output relationships during the sensitivity analysis phase needed to identify key drivers of output uncertainty (Mishra & Knowlton 2003).

The second step, i.e. characterization of uncertain inputs using probability distributions, is a key element in producing a defensible uncertainty analysis study. Unfortunately, a systematic approach to distribution assignment often appears to be ignored in the hydrologic modeling literature. Such a methodology is described in detail by Mishra (2002) and involves the following components:

- Fitting distributions to measured data using probability plotting, non-linear regression techniques or maximum likelihood estimation (e.g. Hahn & Shapiro 1967).
- Deriving distributions using known constraints and the principle of maximum entropy (e.g. Harr 1987), which

forces the analyst to be maximally uncertain with respect to unknown information.

- Assessing subjective distributions using formal expert elicitation protocols (e.g. Keeny & von Winterfeld 1991). A key feature of such protocols is the assessment of selected values (e.g., min, max, median, 10th–90th percentiles, 25th–75th percentiles) and their bases rather than parameters of named distributions (e.g. normal).
- Combining prior information, such as expert elicitation, with new data, such as field measurements using Bayesian updating (e.g. Benjamin & Cornell 1970).

The third step, i.e. parameter sampling, requires selecting an appropriate sampling scheme. The two common choices are purely random sampling (often erroneously referred to as Monte Carlo sampling) and Latin Hypercube sampling (LHS). As the name implies, purely random sampling is a straightforward process and is described in detail in standard references (e.g. Tung & Yen 2005). LHS, which was originally developed for nuclear waste disposal risk analysis applications (McKay *et al.* 1979), is less well known in the hydrologic literature. It is a stratified sampling procedure which involves dividing the range for each input into strata of equal probability, picking one value from each interval and randomly combining values picked for different variables.

Compared to purely random sampling, LHS provides more uniform space-filling coverage of the parameter space. LHS also results in faster convergence for mean and standard deviation of the output—especially if the number of samples is small (as is likely to be the case with many computationally demanding hydrologic models). The restricted pairing technique (Iman & Conover 1982), when used in conjunction with LHS, allows preserving any specified correlation structure between uncertain inputs as well as eliminating spurious input–input correlations.

The fourth step, i.e. multiple model computations, first involves ensuring that enough simulations have been performed to obtain a stable solution. One widely used rule of thumb for selecting an optimal sample size in LHS in order to obtain a stable mean is the $(4/3)N$ rule (Iman & Helton 1985), where N is the number of uncertain inputs. However, additional runs may be required if tail

percentiles are to be used as the performance metrics of interest. A graphical comparison of output CDFs using multiple sample sizes, supplemented by the Kolmogorov–Smirnov test (e.g. Press *et al.* 1992), can also be utilized to evaluate the adequacy of using a smaller sample size.

Once an appropriate sample size has been selected, parametric and non-parametric estimates of the reliability in the model output can be determined. For example, error bars on the expected value of the output can be approximated using the central limit theorem (e.g. Benjamin & Cornell 1970) or simulated via bootstrap techniques (e.g. Tung & Yen 2005). Replicate sample sets generated using different random seeds are sometimes used to provide uncertain bounds on model output statistics (Helton 1993).

MCS results are generally presented in the form of CDFs and/or histograms. For time-dependent models, the CDF can be extracted at different time slices. Uncertainty in the time history can also be presented as running 95% and 5% confidence intervals around the mean or median time-dependent outcome of interest. Box plots are also a popular graphical tool for displaying model results, especially when comparing the behavior of different models and/or modeling scenarios. Other common output statistics include the value at a given probability level (e.g. 95th percentile) or the exceedance probability associated with a target outcome (e.g. risk level of 10^{-6}).

As noted earlier, MCS provides the greatest flexibility for uncertainty propagation so long as data are available to properly define the probability distributions of uncertain inputs, and the model can be run multiple times (e.g. ~ 100 – $1,000$) to ensure stable estimates of the output distribution. In practice, these two conditions may not always be met because: (1) paucity of data may force the analyst to make simplifying assumptions regarding input distributions and (2) computational constraints may preclude performing more than a few tens of forward simulations, especially for distributed parameter hydrologic models. Also, MCS could be an “overkill” solution if the probabilities associated with only a few outcomes of interest are required. Various alternatives to MCS for such situations are described next.

First-order second-moment method (FOSM)

Often, it is sufficient to quantify the uncertainty in model predictions in terms of the mean (describing the central tendency of the prediction) and the variance (describing the spread around the mean) rather than the full distribution. The first-order second-moment method (FOSM) is one such methodology (Benjamin & Cornell 1970; Dettinger & Wilson 1981; Ang & Tang 1984; Morgan & Henrion 1990; Tung & Yen 2005).

The first-order estimate of the expected value (mean) of an uncertain quantity, F , is obtained by using the mean (expected value) of each of the uncertain parameters to evaluate the model, viz:

$$E[F] \cong F(\hat{\mathbf{x}}) \quad (1)$$

where $\hat{\mathbf{x}}$ is the vector of mean values of the uncertain parameters and $E[\cdot]$ denotes the expectation operator. Equation (1) assumes small and uniformly distributed parameter perturbations around the mean values. Similarly, the variance, $V[\cdot]$, is given by

$$V(F) \cong \sum_i \sum_j \left(\frac{\partial F}{\partial x_i} \right) \left(\frac{\partial F}{\partial x_j} \right) C[x_i x_j] \quad (2)$$

where the covariance, $C[x_i x_j] = \rho_{ij} \sigma[x_i] \sigma[x_j]$, is expressed in terms of the parameter correlation coefficients, ρ_{ij} , and the individual parameter standard deviations, σ . The variance of F is thus seen to depend on the variance-covariance relation of the input parameters, and sensitivity of the output to the uncertain inputs. For uncorrelated parameters, the expression for variance simplifies to

$$V(F) \cong \sum_i \left(\frac{\partial F}{\partial x_i} \right)^2 V[x_i] \quad (3)$$

Each term in the summation in Equation (3) can be interpreted as the fractional contribution to total output variance, thus yielding information about the relative importance of each uncertain input.

From a practical standpoint, the first-order estimate of the mean is a reasonable approximation so long as parameter variances are small and the function is only mildly nonlinear. The computation of the variance requires estimation of input-output sensitivity derivatives, for which a finite-difference approximation can be obtained

numerically using $N + 1$ model runs. It is also advisable to perform variable transformations as needed so that the input-output relationship is (quasi) linear and parameter uncertainties are approximately symmetric around the mean values.

In summary, the FOSM technique is an appealing alternative to MCS when only the mean and variance of model outputs are of interest rather than the full CDF. It involves considerably less computational effort for problems with a small number of uncertain parameters, while providing results of comparable accuracy for linear and mildly nonlinear problems (Mishra & Parker 1989; James & Oldenberg 1997).

Point estimate method (PEM)

Although the FOSM technique is conceptual simple, its practical applicability is limited for nonlinear models or for models where numerical computation of derivatives could be burdensome. To overcome these limitations of FOSM and to provide an efficient method for relating the statistical moments of the inputs to the moments of the output, the point estimate method (PEM) was proposed by Harr (1989). In this method, the model is evaluated at a discrete set of points in the uncertain parameter space, with the mean and variance of model predictions computed using weighted averages of these functional evaluations.

The starting point in PEM is the estimation of the eigenvalues (λ_i) and eigenvectors (e_{ij}) of the correlation matrix for the uncertain variables. Each variable, x_j , is then perturbed around its mean by a factor, Δx_j :

$$\Delta x_j = \pm e_{ij} \sqrt{N} \sigma[x_j] \quad (4)$$

where N is the number of uncertain variables and σ denotes the standard deviation. The method thus results in $2N$ point estimates of the model, based on which the output mean is computed as follows:

$$E[F] = \sum_i (F_i^+ + F_i^-) \frac{\lambda_i}{2N} \quad (5)$$

and the output variance is computed from

$$E[F^2] = \sum_i \left[(F_i^+)^2 + (F_i^-)^2 \right] \frac{\lambda_i}{2N} \quad (6)$$

by noting the relationship: $V[F] = E[F^2] - (E[F])^2$. Here F_i^+ and F_i^- denote estimates of model output corresponding to the perturbation of each input parameter from its mean value by Δx_j in the positive and negative directions, and λ_i are the eigenvalues corresponding to each input.

Although $2N$ model evaluations are required to compute the mean and variance as per Equations (5) and (6), it has been noted that in many cases the eigentransformation of the correlation matrix results in only a few dominant eigenvalues (Harr 1989). Thus, it is possible to use this subset of eigenvalues for uncertainty propagation without any significant loss of accuracy.

In summary, the PEM approach is a derivative-free alternative to FOSM for estimating the mean and variance of uncertain model outputs (see Figure 1). The original PEM algorithm of Harr (1989) was designed for correlated random variables with normal distributions. Chang *et al.* (1997) describe a methodology for extending this method to problems involving multivariate non-normal random variables. Unlu *et al.* (1995) and Mishra (2000) provide comparative assessments of FOSM and PEM for uncertainty propagation using subsurface flow and transport models.

Logic tree analysis (LTA)

Logic tree analysis (LTA) is particularly useful for uncertainty propagation when parameter uncertainty is described using a limited number of probable states (e.g. high, medium, low values) and their likelihoods. Logic trees (also known as probability trees) combine individual

scenarios resulting from uncertain discrete events and/or parameter states. As such, they may be recognized as a special case of decision trees containing only chance nodes but no decision nodes (Morgan & Henrion 1990).

The logic tree is ordered such that independent effects are placed to the upstream (left) side and dependent effects are organized to the downstream (right) side. Each branch is assigned a probability, conditional on the values of the previous branches leading to that node. All possibilities must be considered in building the logic tree, such that probabilities for branches originating from each node sum to 1.

Consider a simple groundwater contaminant transport modeling problem involving two uncertain inputs—source concentration (s) and groundwater velocity (v). Uncertainty in the source node is represented by two values, s_1 and s_2 , with probabilities P_1 and P_2 , respectively. Uncertainty in the velocity node is also represented by two values, v_1 and v_2 . These values have conditional probabilities ranging from P_3 to P_6 , depending on which branch of the source node they are attached to. Each path from the root to an end branch (or terminal node) of the tree represents a feasible scenario. The four feasible scenarios for this system can be enumerated as: (s_1, v_1) , (s_1, v_2) , (s_2, v_1) and (s_2, v_2) . The probability of each scenario is the product of conditional probabilities of the branches along that path.

The logic tree thus organizes various parameter combinations and their probabilities. Given this information, the computation of the consequence for each of the discrete combinations is a straightforward task. The results can be organized in terms of a table or graph of sorted discrete outcomes versus the corresponding summed probabilities. Such a “risk profile” is equivalent to a cumulative distribution of model output generated via MCS.

In summary, the LTA methodology is a useful alternative to MCS when uncertainty characterization is based on a limited number of possibilities (as in the case of expert elicitation). Given the combinatorial nature of the algorithm, it can only handle a limited number of uncertain inputs and is often useful in screening-type analyses. An example application of the LTA methodology is in risk assessments of the potential nuclear waste repository at Yucca Mountain, NV (Kessler & McGuire 1999).

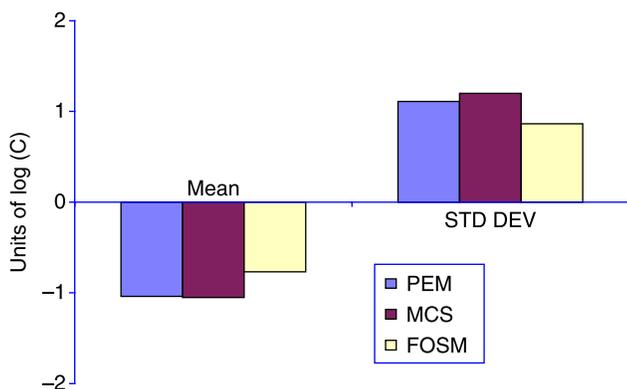


Figure 1 | Comparison of FOSM, PEM and MCS results, from Mishra (2000).

First-order reliability method (FORM)

Although MCS can produce the full distribution of model output, such detailed information may not be necessary when the probability associated with only a limited number of “target” states (e.g. regulatory concentration limit) is of interest. An alternative to MCS for such purposes is the first-order reliability method (FORM), originally developed in the field of structural reliability (e.g. Ang & Tang 1984; Madsen *et al.* 1986).

In FORM, the goal is to evaluate the probability of the outcome of interest, F , exceeding some prescribed target, in terms of the limit-state function, $g[\cdot]$:

$$g(\mathbf{x}) = F_{\text{target}} - F(\mathbf{x}) \quad (7)$$

where \mathbf{x} is the vector of input parameters in the model, $F(\mathbf{x})$. The FORM solution involves first transforming the elements of \mathbf{x} into a set of uncorrelated, standardized normal variables, \mathbf{U} . The limit-state surface, $g(\mathbf{U}) = 0$, is then approximated by a tangent hyperplane at the point closest to the origin, which is also known as the “most probable point” or “design point”. The coordinates of this point are obtained by solving the nonlinear minimization problem:

$$\min \beta = \sqrt{\mathbf{U}^T \mathbf{U}}; \quad \text{s.t. } g(\mathbf{U}) = 0 \quad (8)$$

which leads to a first-order approximation of the exceedance probability, P^* , as

$$P^* \cong \Phi(-\beta) \quad (9)$$

where $\Phi(\cdot)$ is the standard normal CDF and β , also known as the reliability index, is the distance from the origin to the approximating tangent hyperplane in the transformed \mathbf{U} -space. The reliability index can also be conceptualized as the separation between the origin and the limit-state surface measured in units of standard deviation (e.g. Ang & Tang 1984).

In practical applications of FORM, one specifies a target outcome and computes the associated exceedance probability. This is equivalent to locating one point on the cumulative distribution function (CDF), i.e. a graph of model outcome, F , vs. cumulative probability, P , where $P = 1 - P^*$. Repeating this process for other “target states” provides multiple pairs of outcome and cumulative prob-

ability values, from which the full CDF of model outcome can be constructed and compared against conventional Monte Carlo simulation results.

In summary, FORM can be an efficient alternative to MCS or LTA for computing the probability of a limited number of outcomes of interest (see Figure 2). However, when the system model is nonlinear and/or uncertain inputs have large variance, higher-order approximations to FORM may be required to obtain accurate estimates of exceedance probabilities (Madsen *et al.* 1986). In such cases, reliability-based methods provide no particular advantage over MCS. FORM has been applied for uncertainty analyses in a variety of hydrological problems (Sitar *et al.* 1987; Melching 1992; Hamed *et al.* 1995; Xiang & Mishra 1997).

SENSITIVITY ANALYSIS TECHNIQUES

Traditional sensitivity analysis of distributed parameter hydrologic models involves perturbing each of the parameters by a small amount, one at a time, from a reference value and computing the corresponding change in the model output. Sensitivity coefficients, computed as the change in output divided by the change in input, reflect the slope of the input–output relationship at the reference point. However, unless the functional relationship between the output and the input of interest is linear over the entire range of input values, such analyses can only provide information regarding the relative sensitivities of input parameters that is valid locally. The “one-off” nature of

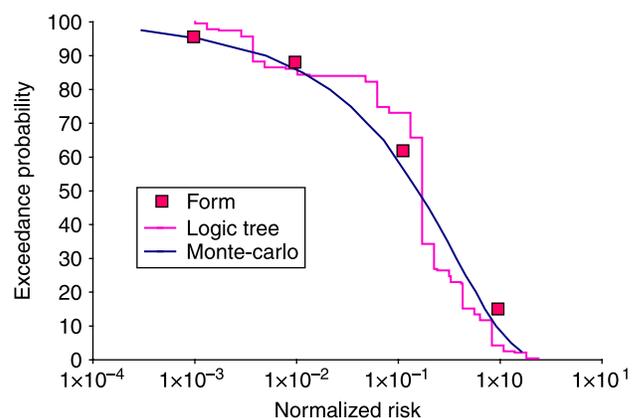


Figure 2 | Comparison of LTA, FORM and MCS results, from Mishra & Freeze (1998).

these analyses also precludes a proper accounting for synergistic effects between model inputs.

In the context of uncertainty analysis, it is more useful to apply global sensitivity analysis techniques that can investigate input–output sensitivities valid over the full range of parameter variations and combinations (Saltelli *et al.* 2000). Such techniques are useful for examining the sensitivities of model results (and their uncertainties) to the uncertainties and assumptions in model inputs. Global sensitivity analysis, also referred to as uncertainty importance analysis, generally involves examining probabilistic model results for multiple goals, viz: (a) to identify key contributors to output uncertainty, (b) to determine key factors controlling extreme model outcomes or (c) to detect the presence of non-monotonic input–output patterns (Mishra *et al.* 2002).

In this section, various sensitivity (uncertainty importance) analysis techniques are described and their advantages and limitations are compared. The methods selected are those that can be used in conjunction with an existing sampling-based uncertainty analysis. As such, they are most appropriate for computation-intensive hydrologic models, i.e. where re-sampling of the uncertain parameters and re-computation of model results is not practical.

It is useful to note that the contribution to output uncertainty (variance) by an input is a function of: (a) uncertainty of the input variable and (b) sensitivity of the output to that particular input. In general, input variables identified as important in global sensitivity analysis have both characteristics: they demonstrate significant variance and are characterized by large sensitivity coefficients. Conversely, variables which do not show up as important per these metrics are either restricted to a small range in the probabilistic analysis, and/or exhibit limited input–output sensitivity.

Stepwise regression analysis

A popular framework for uncertainty importance analysis involves building a multivariate linear rank regression model of the form

$$\hat{y} = b_0 + \sum_j b_j x_j \quad (10)$$

where \hat{y} denotes the (predicted) rank-transformed output, the x_j are the rank-transformed input variables of interest and the b_j are the unknown coefficients (Helton 1993). Rank transformation, where variables are ranked in ascending order with their values replaced by the ranks, is the simplest non-parametric linearizing technique (Iman & Helton 1985).

The regression coefficients are generally determined using a stepwise regression procedure (Draper & Smith 1981). A sequence of regression models is constructed starting with the parameter that explains the largest amount of variance in the output. At each successive step, the parameter that explains the largest fraction of residual variance is included in the model. The process continues until all input variables that explain statistically significant amounts of variance in the output have been included in the model.

A commonly used measure of uncertainty importance is standardized regression coefficient (SRC), defined as

$$\text{SRC} = \frac{b_j \sigma(x_j)}{\sigma(y)} \quad (11)$$

where $\sigma(x_j)$ denotes the sample standard deviation of the uncertain input x_j and $\sigma(y)$ denotes the sample standard deviation of the output y . Thus, the SRCs can be interpreted as regression coefficients that would be obtained from a regression analysis with the input and output variables normalized to zero mean and unit standard deviation. The square of the SRC can also be shown to be equal to the fractional contribution to variance for uncorrelated inputs (Iman & Helton 1985).

When variables are correlated, a more appropriate measure of uncertainty importance is the partial correlation coefficient (PCC). PCCs quantify the strength of a linear relationship between input–output pairs after eliminating the linear influence of other input variables (Draper & Smith 1981). They can be readily calculated from the input–input correlation matrix and the input–output correlation vector using simple matrix algebra. RamaRao *et al.* (1998) showed that the square of the PCC can be interpreted as the gain in R^2 of an input–output regression model—when the selected variable is brought into regression—as a fraction of the currently unexplained variance.

In summary, input–output regression modeling is a powerful tool for identifying key contributors to output variance. The methodology is limited to monotonic and (rank-transformed) linear patterns of association. Example hydrologic applications of this technique include Gwo *et al.* (1996), who applied it to a subsurface stormflow model, and Muleta & Nicklow (2005), who applied it to a watershed model.

Entropy (mutual information) analysis

Since the concept of correlation and regression is strictly applicable to monotonic relationships (RamaRao *et al.* 1998), it is useful to pose the sensitivity analysis problem in the general terms of identifying important non-random patterns of association. Here, the word “association” is used in a broader context than “correlation” and includes both monotonic and non-monotonic relationships. Determining the significance and strength of input–output association is facilitated by the information-theoretic concept of entropy, which provides a useful framework for the characterization of uncertainty (or information) in the univariate case and redundancy (or mutual information) in the multivariate case (e.g. Press *et al.* 1992). Mishra & Knowlton (2003) describe a methodology for global sensitivity analysis that combines the mutual information concept with contingency table analysis.

As per Press *et al.* (1992), let the input variable x have I possible states (labeled by i) and the output variable y have J possible states (labeled by j). This information can be compactly organized in terms of a contingency table, a table whose rows are labeled by the values of the independent variable, x , and whose columns are labeled by the values of the dependent variable, y . The entries of the contingency table are non-negative integers giving the number of observed outcomes for each combination of row and column. The corresponding probabilities are readily obtained by normalization.

The mutual information between x and y , which measures the reduction in uncertainty of y due to knowledge of x (or vice versa), is defined as (e.g. Bonnlander & Weigend 1994)

$$I(x, y) = \sum_i \sum_j p_{ij} \ln \frac{p_{ij}}{p_i \cdot p_j} \quad (12)$$

Here p_{ij} is the probability of outcomes corresponding to both state x_i and state y_j , while p_i is the probability of outcomes corresponding to state x_i alone and p_j is the probability of outcomes corresponding to state y_j alone. A useful measure of importance defined on the basis of mutual information is the so-called R -statistic (Granger & Lin 1994):

$$R(x, y) = [1 - \exp\{-2I(x, y)\}]^{1/2} \quad (13)$$

where R takes values in the range [0,1], with values increasing with I . R is zero if x and y are independent, and is unity if there is an exact linear or nonlinear relationship between x and y .

In summary, the entropy-based measure R -statistic is a very general tool for quantifying the strength of an association, even in non-linear and/or non-monotonic cases. Mishra & Knowlton (2003) show how important non-monotonic patterns, missed by stepwise regression analysis, can be readily identified using entropy analysis (see Figure 3).

Classification tree analysis

Sensitivity analyses based on stepwise regression or mutual information concepts are typically applied to the entire spectrum of input–output data. However, specialized approaches may be required for examining small subsets

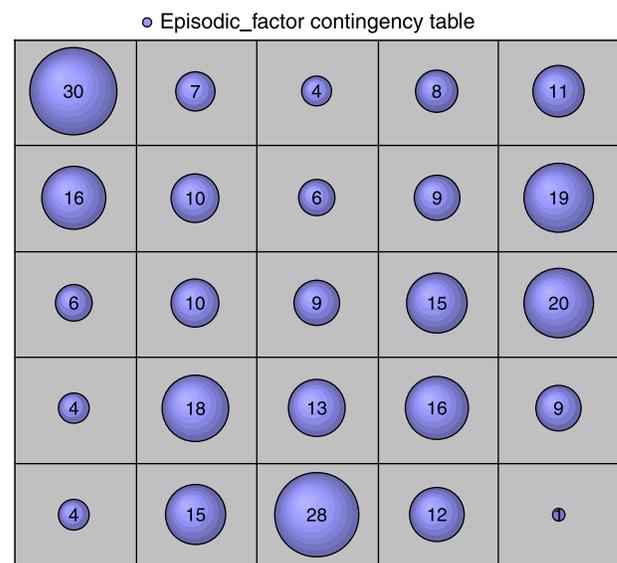


Figure 3 | Bubble plot showing visualization of contingency table for a key non-monotonic input identified from entropy analysis, from Mishra & Knowlton (2003).

(e.g. top and bottom deciles) of the output. To this end, classification tree analysis can provide useful insights into what variable or variables are most important in determining whether outputs fall into one or the other (extreme) category (Breiman *et al.* 1984). Classification tree analysis has been most widely used in medical decision-making. In hydrological modeling, such categorical problems may arise in the context of calibration, where the factors contributing to good vs. poor fits may be of interest. Another example is contaminant transport, where insights on variables responsible for high vs. low migration distance may be useful.

A binary decision tree is at the heart of classification tree analysis. The decision tree is generated by recursively finding the variable splits that best separate the output into groups where a single category dominates. For each successive fork of the binary decision tree, the algorithm searches through the variables one by one to find the purest split within each variable. The splits are then compared among all the variables to find the best split for that fork. The process is repeated until all groups contain a single category. In general, the variables that are chosen by the algorithm for the first several splits are the most important, with less important variables involved in the splitting near the terminal end of the tree.

One commonly used tree-building methodology is based on a probability model approach (Venables & Ripley 1997). Classifiers at each node are selected based on an overall maximum reduction in deviance, for all possible binary splits over all the input variables. The deviance at a given node i is

$$D_i = -2 \sum_k n_{ik} \log(p_{ik}) \quad (14)$$

where p_{ik} is the probability distribution at node i over the classes k , and n_{ik} is the number of cases (y values) assigned to k at i . The probabilities are unknown, but can be estimated from the proportions at each node, i.e. $p_{ik} = n_{ik}/n_i$. The reduction in deviance from splitting node s into nodes t and u would then be

$$D_s - D_t - D_u = 2 \sum_k \left[n_{tk} \log\left(\frac{n_{tk}n_s}{n_{sk}n_t}\right) + n_{uk} \log\left(\frac{n_{uk}n_s}{n_{sk}n_u}\right) \right] \quad (15)$$

The classification tree is built by successively taking the maximum reduction in deviance over all the allowed

splits of the leaves to determine the next split. Termination occurs when the number of cases at a node drops below a set minimum, or when the maximum possible reduction in deviance for splitting a particular node drops below a set minimum.

In summary, classification tree analysis is a powerful tool for determining variable importance for categorical problems (see Figure 4). Compared to linear regression modeling, tree-based models are attractive because: (a) they are adept at capturing non-additive behavior, (b) they can handle more general interactions between predictor variables and (c) they are invariant to monotonic transformations of the input variables. Mishra *et al.* (2003) describe an application of this methodology for identifying key variables affecting extreme outcomes in a groundwater-driven radionuclide transport model. Ruskauff *et al.* (2006) show how useful decision rules regarding calibration of a regional-scale groundwater model can be extracted from classification tree analysis.

EXAMPLE APPLICATIONS

Case study 1: uncertainty in flows and loads from a watershed

In this study, the WaterShed water quality model (WASH) was used to develop a probabilistic model for the C-44 reservoir/stormwater treatment area located in the St. Lucie

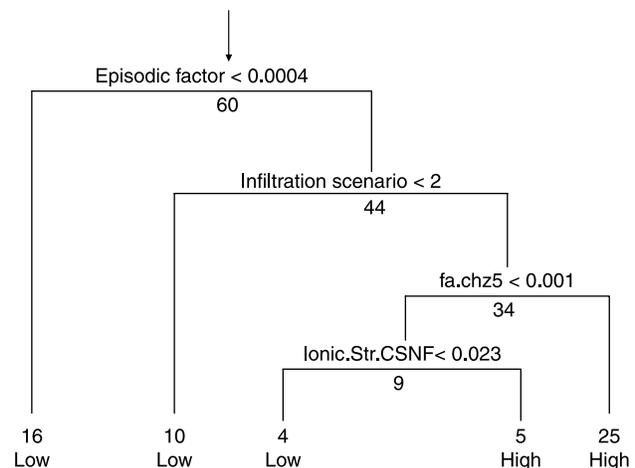


Figure 4 | Decision tree based on classification tree analysis, from Mishra *et al.* (2002).

Estuary Watershed basin in South Florida. WASH performs time-dependent simulations of hydrologic, hydrodynamic and water quality conditions in surface water basins and waterbodies. It is closely related to the Hydrological Simulation Program–Fortran (HSPF) model developed by the USGS and the US EPA. The goal of this study was to provide confidence intervals for flows and loads at the outlet of the watershed and also to identify key input parameters contributing the most to the predicted uncertainty in flows and loads.

Uncertain parameters considered for this study were selected on the basis of a sensitivity analysis conducted as part of the C-44 watershed model development. These include:

- hydrologic parameters CEPSC (interception storage capacity), INFILT (infiltration) and SLSUR (overland flow plane slope);
- irrigation parameters S_CANAL, S_EXT, S_SH_AQ (% irrigation demand withdrawn from each source), and SPRAY, SURFACE, LOWER_Z, UPPER_Z (% irrigation applied to each irrigation target);
- hydrogeologic parameter COND (canal conductance);
- water quality parameters SQOLIM1 (event mean concentration for total nitrogen) and SQOLIM2 (event mean concentration for total phosphorus).

Based on data availability, a variety of techniques was used for uncertainty characterization of these parameters. A log-normal distribution was fitted for CEPSC using 10 data points from the literature. A triangular distribution was assigned to INFILT based on an evaluation of related model studies. A maximum entropy triangular distribution was chosen for SLSUR based on known constraints about local agricultural practice. Maximum entropy uniform distributions were used for both groups of irrigation parameters. An expert elicitation was carried out to characterize the uncertainty in COND. Finally, maximum entropy triangular distributions were chosen for SQOLIMs using a limited number of values from the literature.

250 Latin Hypercube samples of these uncertain inputs were generated and then passed to the WASH process model to compute the total flows and loads for each realization from 1965–2000. For illustration purposes, the variation in total phosphorus (TP) loading for the time

window July–September 1971 is shown in Figure 5. Depicted here are all 250 realizations along with the TP load time series for the mean and 5–95% confidence bounds. In general, the model appears to be well-behaved with respect to the uncertainty bounds.

The CDF expressing the uncertainty in the integrated TP load (i.e. at the end of the 35-yr simulation period) is illustrated in Figure 6. Also shown here is the CDF for a sample size of 100, which shows a less stable response compared to the 250 sample cases. Such simple graphical comparisons can be used to provide confidence that an appropriate sample size was used in the MCS process.

The FOSM methodology was also employed to obtain the uncertainty in the flows and loads at the downstream outlet of the C-44 watershed. Although computationally efficient, the FOSM techniques can only quantify the uncertainty in model predictions in terms of the mean (describing the central tendency of the prediction) and the variance (describing the spread around the mean), rather than the full distribution. Since there were 11 uncertain inputs, a total of 12 WASH model runs was required to develop the FOSM results. The standard deviations for the uncertain inputs were estimated from the distributions prescribed for the MCS application. Figure 7 shows excellent agreement between MCS and FOSM generated mean and standard deviation in TP load—even though the computational burden of FOSM is lower by a factor of ~20.

Sensitivity (uncertainty importance) analysis of the MCS results was conducted to evaluate the relationships

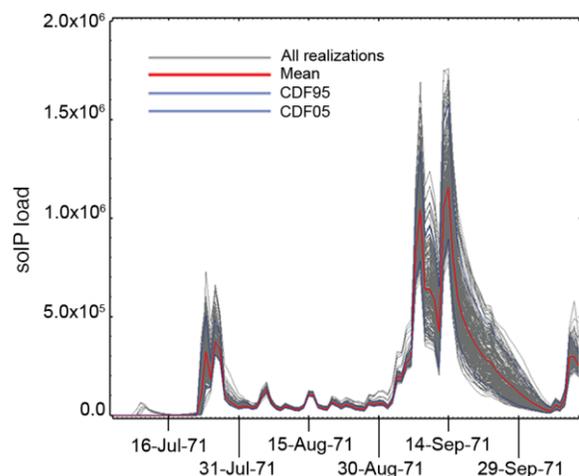


Figure 5 | Uncertainty in time history of total phosphorus load, case study 1.

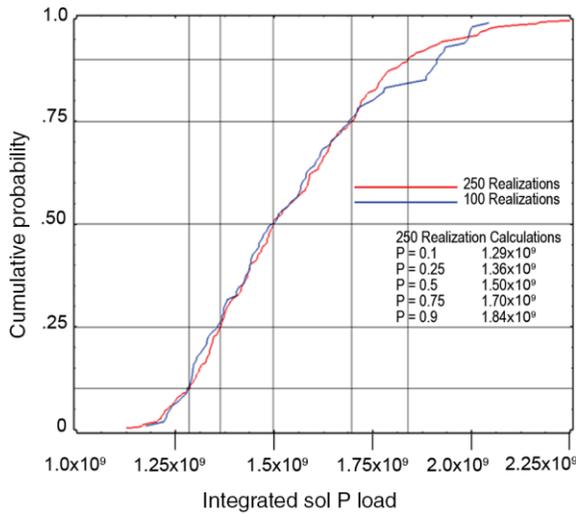


Figure 6 | CDF of total phosphorus load showing sensitivity to sample size, case study 1.

between uncertain model inputs and uncertainty in TP loads. Stepwise regression analysis indicated that the primary contributor to the uncertainty in TP load was the uncertainty in the monthly distribution of event mean phosphorus concentration values (SQOLIM2). The entropy analysis indicated the same variable importance ranking as stepwise regression analysis, indicating the absence of any significant non-monotonic input-output patterns. The strong correlation between TP load and SQOLIM2 is clearly evident in the scatter plot presented in Figure 8. This highlights the need to constrain the range of SQOLIM2 using as much information and/or real data as possible in order to minimize model prediction uncertainty.

The sensitivity analysis results for total phosphorus were confirmed using a conditional analysis where the most important parameter SQOLIM2 was fixed at its median

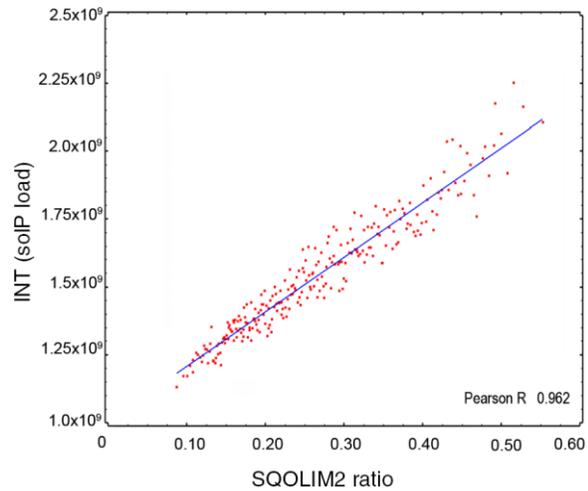


Figure 8 | Scatter plot showing the most important input-output relationship based on stepwise regression, case study 1.

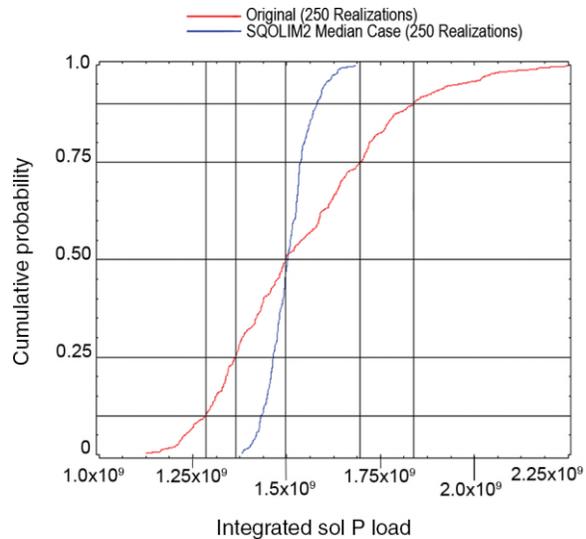


Figure 9 | Conditional CDF showing effect of holding most important uncertain input at median value, case study 1.

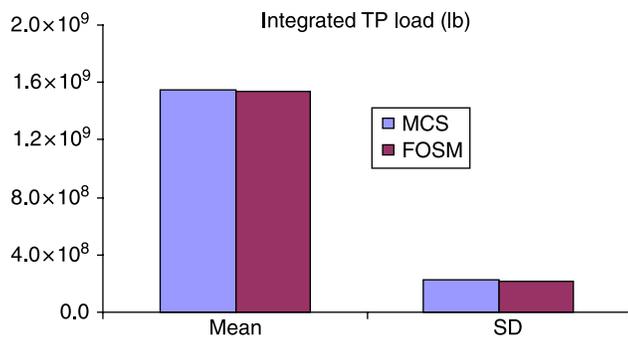


Figure 7 | Comparison of FOSM and MCS results, case study 1.

value. As expected, Figure 9 shows the corresponding reduction in output uncertainty, indicated by the narrowing in the range of the CDF.

In summary, this case study demonstrates how uncertainty and sensitivity analysis techniques can be applied in a systematic manner to field-scale hydrologic models. When the model is reasonably linear and well-behaved, and input uncertainties are not too large, FOSM is a viable alternative to MCS for uncertainty propagation. In the absence of non-monotonic input-output patterns, stepwise regression analysis yields robust rankings of variable importance.

Case study 2: uncertainty in hydrologic outcomes of a regional simulation model

The Regional Simulation Model (RSM), developed by the South Florida Water Management District, is a numerical tool for simulating surface water and groundwater interactions in shallow water table environments. One specific application of the RSM is the Natural Systems Regional Simulation Model (NSRSM), which was designed to simulate the predevelopment hydrologic response using appropriate land use conditions and topography. The goal of this study was to apply uncertainty and sensitivity analysis techniques to demonstrate quantification of the uncertainty in predicted hydrologic outcomes of the NSRSM.

Uncertain input parameters were taken to be Manning's coefficient, detention storage, vegetation coefficient, extinction depth and storativity for two major land use types. These variables were each prescribed a distribution based on standard approaches discussed earlier, including a review of available literature, model calibration data from analogous regions, known constraints and expert judgment. Topographic uncertainty was characterized using "low", "base" and "high" maps. The output metrics, calculated at two locations, were taken to be: (a) water stage averaged for a one-week time window within a representative year and (b) daily-averaged transect flow.

In the uncertainty analysis phase, cases of 100, 200 and 300 realizations were run with the NSRSM model, and the results of the 200-realizations case were chosen for detailed analysis. Figure 10 shows the uncertainty in the time history of transect flow at one of the locations, along with several of the percentiles. The extreme variability in model behavior (i.e. occasional large fluctuation in flow with time) leads to maximum extremes that are significantly higher than the 95th percentile outcome and minimum extremes that are significantly lower than the 5th percentile outcome.

In addition to the uncertainty in the time history, the uncertainty in transect flow can also be shown in terms of a CDF that reflects variations from realization to realization for one-week time averaging windows, shown as vertical lines in Figure 10. Note that three time slices have been selected in order to examine the sensitivity of the shape of the CDF to the exact placement of the one-week averaging

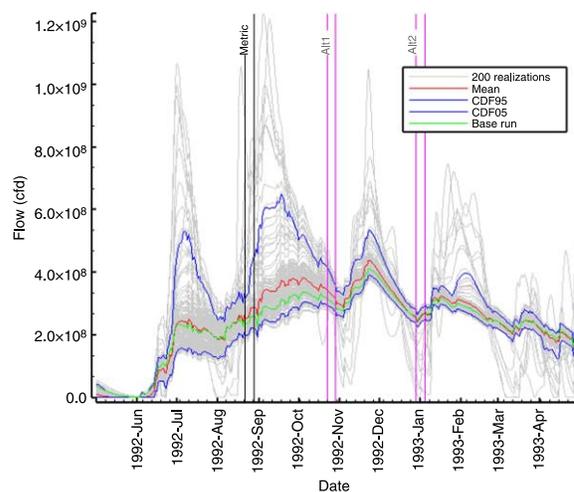


Figure 10 | Uncertainty in time history of transect flow, case study 2.

window. These are denoted as [Metric], [Alt1] and [Alt2], with the corresponding CDFs presented in Figure 11.

The reference CDF (denoted as [Metric]) has considerable right skew, because of the presence of outliers above the 95th percentile bound, as shown in the time history in Figure 10. Conversely, the [Alt2] CDF has considerable left skew because of the presence of outliers below the 5th percentile bound. The [Alt1] CDF is more symmetric, because the outliers are in close proximity to the 5th/95th-percentile bounds. Thus, the degree of skewness seen in the CDF can be directly related to the behavior of the outliers in the time history plot.

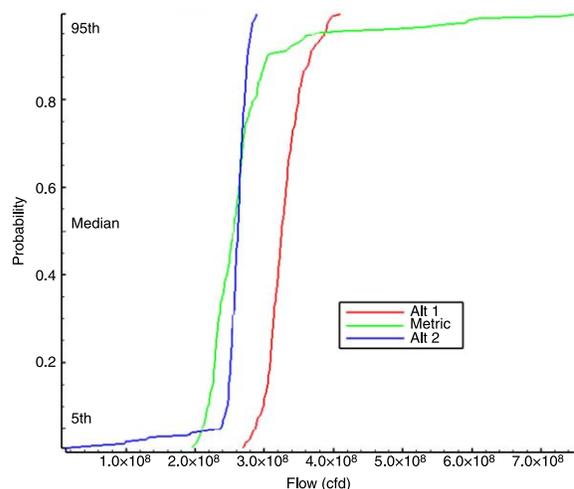


Figure 11 | Sensitivity of CDF to choice of placement of averaging window, case study 2.

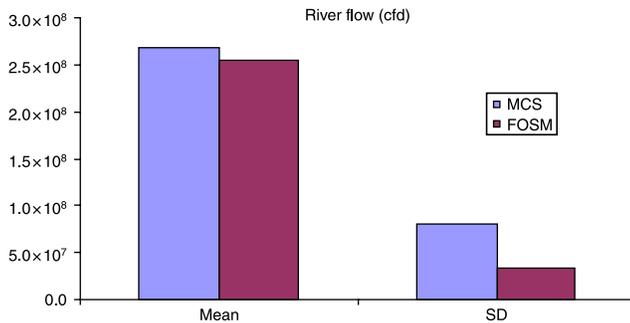


Figure 12 | Comparison of FOSM and MCS results, case study 2.

Output uncertainty was also calculated using FOSM. Recall that numerical derivatives (estimated using a forward-difference approximation) are usually required to calculate the variance in the output. However, single-point derivatives were found to be inaccurate for all output metrics—thus requiring considerably more model simulations to generate stable (multi-point) derivatives. The FOSM analysis was carried out for all of the variables, with the exception of topography, since categorical variables are not amenable to derivative calculations. As shown in Figure 12, the means from FOSM and MCS agree reasonably well, whereas the standard deviation is different by a factor of 2—most likely due to the skewness in the output distribution.

With respect to sensitivity analysis, the output uncertainty was found to be dominated by a single variable, vegetation coefficient. This result was consistent across both classification tree and stepwise regression analysis, and will not be discussed further.

In summary, this case study demonstrates how uncertainty analysis can be challenging in some hydrologic models, primarily due to the nonlinear and discontinuous nature of the output metrics of interest. It also points to the need for verifying the stability of numerical derivatives prior to applying FOSM—especially if FOSM is the only technique chosen for uncertainty propagation.

SUMMARY AND CONCLUDING REMARKS

Uncertainty should be routinely incorporated in the analyses of hydrologic systems to improve the quality of

engineering inputs for decision-making purposes. To this end, it is desirable to adopt a systematic framework for uncertainty and sensitivity analysis that is rigorously conditioned to the quality and quantity of available data, and the objectives of the study.

For uncertainty analysis, MCS is the strategy of choice because it is more general and requires fewer assumptions than other methods discussed in this paper—although it can be more data- and computation-intensive. Recommended steps for a proper MCS study include: (a) assigning distributions based on data, or using appropriate alternatives as discussed earlier, (b) adopting a proper sampling scheme with correlation control to ensure uniform coverage of the parameter space and (c) checking for stability and reliability in model outcomes.

Once the sampling-based uncertainty analysis is complete, sensitivity analysis should be used for examining how model results depend on the uncertainties in model inputs. Some of the techniques for this purpose are: (a) regression analysis to determine the most important contributors to the spread in probabilistic model results, (b) classification tree analysis to identify those variables controlling extreme outcomes in the full suite of probabilistic results and (c) entropy analysis to quantify the strength of input–output association for non-monotonic patterns.

Alternatives to MCS for uncertainty propagation include: (a) FOSM and PEM for computing the mean and standard deviation in model output given the mean, standard deviation and correlation matrix of uncertain inputs, (b) LTA for computing the risk profile given a discrete probabilistic representation of the uncertain inputs and (c) FORM for computing the probability associated with a limited number of target states given the same information as FOSM or PEM. Depending on the goals of the study and the availability of data, one or more of these techniques may be appropriate.

The two case studies presented in this paper indicate the differences in response between a “well-behaved” model and a “challenging” model. In particular, they highlight the need for cross-checking probabilistic model outputs when the model response exhibits discontinuity or large fluctuations in its response. In other words, it is useful to ensure that there is minimal uncertainty about the uncertainty analysis results.

Although the focus of this paper has been parametric uncertainty, some of the recent work in propagation of model uncertainties is worth mentioning. These include Generalized Likelihood Uncertainty Estimation (Beven & Binley 1992), Maximum Likelihood Bayesian Model Averaging (Neuman 2003) and Multi-Model Inference (Poeter & Anderson 2005). Such techniques assess the likelihood of each model using some goodness-of-fit metric between observations and model predictions. Treatment of model uncertainty is likely to be an active area of research in the coming years.

In conclusion, ignoring the ubiquity of uncertainty or using *ad hoc* approaches in modeling natural systems can only compromise the quality of engineering decision-making. As Pliny the Elder has observed, “The only certainty is uncertainty”.

ACKNOWLEDGEMENTS

The contribution of my colleagues, Neil Deeds (INTERA) and Banda RamaRao (Framatome AREVA), towards the development and refinement of several of the ideas, techniques and applications discussed in this paper is gratefully acknowledged.

REFERENCES

- Ang, A. H.-S. & Tang, W. H. 1984 *Probability Concepts in Engineering Planning and Design, Vol. II Decision, Risk and Reliability*. Wiley, New York.
- Benjamin, J. R. & Cornell, C. A. 1970 *Probability, Statistics and Decisions for Civil Engineers*. McGraw Hill, New York.
- Beven, K. J. & Binley, A. 1992 The future of distributed models: model calibration and uncertainty prediction. *Hydrol. Process.* **6**, 279–298.
- Bonnlander, B. V. & Weigend, A. S. 1994 Selecting input variables using mutual information and nonparametric density estimation. In: *Proceedings of the International Symposium on Artificial Neural Networks (ISANN'94)*, Tainan, Taiwan, pp. 42–50.
- Breiman, L., Friedman, J. H., Olshen, R. A. & Stone, C. J. 1984 *Classification and Regression Trees*. Wadsworth and Brooks/Cole, Monterey, CA.
- Chang, C.-H., Yang, J.-C. & Tung, Y.-K. 1997 Uncertainty analysis by point estimate methods. *ASCE J. Hydraul. Eng.* **123** (3), 244–250.
- Dettinger, M. D. & Wilson, J. L. 1981 First order analysis of uncertainty in numerical models of groundwater flow. *Water Resour. Res.* **17** (1), 149–157.
- Draper, N. R. & Smith, H. 1981 *Applied Regression Analysis*. Wiley, New York.
- EPA 1997 *Guiding Principles for Monte Carlo Analysis*. EPA/630/R-97/001, Environmental Protection Agency, Washington, DC.
- Granger, C. W. J. & Lin, J. 1994 Using mutual information to identify lags in nonlinear models. *J. Time Ser. Anal.* **15**, 371–384.
- Gwo, J. P., Toran, L. E., Morris, M. D. & Wilson, G. V. 1996 Subsurface stormflow modeling with sensitivity analysis using a Latin-hypercube sampling technique. *Ground Water* **34** (5), 811–818.
- Hahn, G. J. & Shapiro, S. S. 1967 *Statistical Models in Engineering*. Wiley, New York.
- Hamed, M. M., Conte, J. P. & Bedient, P. B. 1995 Probabilistic screening tool for groundwater contamination and risk assessment. *ASCE J. Environ. Eng.* **121** (11), 767–774.
- Harr, M. E. 1987 *Reliability-Based Design in Civil Engineering*. McGraw-Hill, New York.
- Harr, M. E. 1989 Probabilistic estimates for multivariate analyses. *Appl. Math. Model.* **13** (5), 313–318.
- Helton, J. C. 1993 Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal. *Reliab. Eng. Syst. Saf.* **42**, 327–373.
- Iman, R. L. & Conover, W. J. 1982 A distribution free approach to inducing rank correlation among inputs. *Commun. Stat. Simul. Comput.* **11**, 335–360.
- Iman, R. L. & Helton, J. C. 1985 Investigation of uncertainty and sensitivity analysis techniques for computer models. *Risk Anal.* **8** (1), 71–90.
- James, A. L. & Oldenburg, C. M. 1997 Linear and Monte Carlo uncertainty analysis for subsurface multiphase contaminant transport. *Water Resour. Res.* **33** (11), 2495–2503.
- Keeny, R. L. & von Winterfeld, D. 1991 Eliciting probabilities from experts in complex technical problems. *IEEE Trans. Eng. Manage.* **38** (3), 191–201.
- Kessler, J. H. & McGuire, R. M. 1999 Total system performance assessment using a logic tree approach. *Risk Anal.* **19**, 915–932.
- Madsen, H., Krenk, S. & Lind, N. C. 1986 *Methods of Structural Safety*. Prentice-Hall, Englewood Cliffs, NJ.
- McKay, M. D., Conover, W. J. & Beckman, R. J. 1979 A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* **21** (3), 239–245.
- Melching, C. S. 1992 An improved first-order reliability approach for assessing uncertainties in hydrologic modeling. *J. Hydrol.* **132**, 157–177.
- Mishra, S. 2000 Uncertainty propagation using the point estimate method. In *Calibration and Reliability in Groundwater Modeling: Coping with Uncertainty* (ed. F. Stauffer, W. Kinzelbach, K. Kovar & E. Hoehn), pp. 292–296. IAHS Publ. 265, International Association of Hydrological Sciences. Wallingford, UK.
- Mishra, S. 2002 *Assigning Probability Distributions to Input Parameters of Performance Assessment Models*. Report SKB-TR-02-11. Swedish Nuclear Fuel and Waste Management Co., Stockholm.

- Mishra, S. 2003 Beyond Monte Carlo simulation—alternative techniques for probabilistic groundwater modeling. In *Groundwater Quality Modeling and Management under Uncertainty* (ed. S. Mishra), pp. 204–214. American Society of Civil Engineers, Reston, VA.
- Mishra, S., Deeds, N. E. & RamaRao, B. S. 2003 Application of classification trees in the sensitivity analysis of probabilistic model results. *Reliab. Eng. Syst. Saf.* **73**, 123–129.
- Mishra, S. & Freeze, G. A. 1998 *Probabilistic Risk Assessment Techniques Applied to the Performance Assessment of the Disposal of High-Level Radioactive Waste in Japan*. Report prepared for Advanced Information Technology, Inc. Tokyo, Japan.
- Mishra, S. & Knowlton, R. G. 2003 Testing for input-output dependence in performance assessment models. In: *Proceedings of the Tenth International High-Level Radioactive Waste Management Conference, Las Vegas, NV*. American Nuclear Society, Illinois, USA, pp. 882–887.
- Mishra, S. & Parker, J. C. 1989 Effects of parameter uncertainty on predictions of unsaturated flow. *J. Hydrol.* **108**, 19–25.
- Mishra, S., Statham, W. H., Deeds, N. E. & RamaRao, B. S. 2002 Measures of uncertainty importance for Monte Carlo simulation results. In: *Proceedings of the International Topical Meeting on Probabilistic Safety Assessment (PSA'02), Detroit, MI*. American Nuclear Society, Illinois, USA, pp. 435–442.
- Morgan, M. G. & Henrion, M. 1990 *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. Cambridge University Press, Cambridge.
- Morris, M. D. 1991 Factorial sampling plans for preliminary computational experiments. *Technometrics* **33** (2), 161–174.
- Muleta, M. K. & Nicklow, J. W. 2005 Sensitivity and uncertainty analysis coupled with automatic calibration for a distributed watershed model. *J. Hydrol.* **306**, 127–145.
- Neuman, S. P. 2003 Maximum likelihood Bayesian averaging of uncertain model predictions. *Stochast. Environ. Res. Risk Assess.* **17** (5), 291–305.
- Poeter, E. & Anderson, D. 2005 Multimodel ranking and inference in ground water modeling. *Ground Water* **43** (4), 597–605.
- Press, W. H., Teukloksky, S. A., Vetterling, W. T. & Flannery, B. P. 1992 *Numerical Recipes in Fortran*. Cambridge University Press, Cambridge.
- RamaRao, B. S., Mishra, S. & Andrews, R. W. 1998 Uncertainty importance of correlated variables in a probabilistic performance assessment. In: *Proceedings of the SAMO'98, Second International Symposium on Sensitivity Analysis for Model Output, Venice, Italy, 19–22 April*. European Commission, Luxembourg, pp. 215–218.
- Ruskauff, G. J., Mishra, S., Ewing, J. E. & Deeds, N. E. 2006 Local and global sensitivity analysis for a complex flow model. In: *Proceedings, MODFLOW and More 2006, Golden, CO*.
- Saltelli, A., Chan, K. & Scott, M. (Eds.) 2000 *Sensitivity Analysis*. Wiley, Chichester
- Sitar, N., Cawfield, J. D. & der Kiureghian, A. 1987 First-order reliability approach to stochastic analysis of subsurface flow and solute transport. *Water Resour. Res.* **23** (5), 794–804.
- Tung, Y.-K. & Yen, B.-C. 2005 *Hydrosystems Engineering Uncertainty Analysis*. McGraw-Hill, New York.
- Unlu, K., Parker, J. C. & Chong, P. K. 1995 Comparison of three uncertainty analysis methods to assess impacts on groundwater of constituents leached from land-disposed waste. *Hydrogeol. J.* **3** (2), 4–18.
- Venables, W. N. & Ripley, B. D. 1997 *Modern Applied Statistics with S-PLUS*. Springer-Verlag, New York.
- Xiang, Y. & Mishra, S. 1997 Probabilistic multiphase flow modeling using the limit-state method. *Ground Water* **35** (5), 820–826.

First received 14 May 2008; accepted in revised form 24 July 2008