International Conference "Water, Environment, Energy and Society" (WEES–2009)
New Delhi, 12–16 January 2009

Predictive Uncertainty in Hydrologic and Water Quality Modeling: Approaches, Application to Environmental Management, and Future Challenges

Mohamed M. Hantush[1]
Land Remediation and Pollution Control Division, National Risk Management Research Laboratory,
Office of Research and Development, U.S. Environmental Protection Agency Cincinnati, OH - 45069, USA
E-mail: Hantush.mohamed@epa.gov

Latif Kalin
School of Forestry and Wildlife Sciences
Auburn University, Auburn, AL - 36849, USA
E-mail: latif@auburn.edu

Rao S. Govindaraju
School of Civil Engineering
Purdue University, West Lafayette, IN - 47907, USA
E-mail: govind@ecn.purdue.edu

ABSTRACT: Extant process-based hydrologic and water quality models are indispensable to water resources planning and environmental management. However, models are only approximations of real systems and often calibrated with incomplete and uncertain data. Reliable estimates, or perhaps further, reduction of prediction uncertainty, contribute directly to successful risk management and the formulation of environmental policy. This paper discusses widely used and promising methods for estimating model prediction uncertainty in complex environmental systems; and lays down a framework for probabilistic risk management and its application to one of the most important watershed-based regulatory programs, the Total Maximum Daily Loading (TMDL). In TMDL development, a Margin of Safety (MOS) is applied to account for the uncertainty embedded in the analysis or modeling exercise. However, in most TMDL developments, MOS is arbitrarily selected and the related degree of protection provided by the safety factor often remains unknown. A formal risk-based approach linking required load reduction in a TMDL to the analysis uncertainty and required degree of protection is presented along with a formal estimation of MOS is also presented. Bayesian-based probabilistic approaches, such as Classical Bayesian Estimation (BEA), Generalized Likelihood Uncertainty Estimation (GLUE), and Ensemble Kalman Filter (EnKF) hold promise for TMDL development under conditions of uncertainty. Current TMDL practices need to be revised taking into consideration recent advances in model uncertainty estimation. The paper ends with a list of future challenges in uncertainty estimation and research needs to reduce its magnitude.

INTRODUCTION

Population growth and urbanization worldwide threaten to undermine water resources availability and degrade water quality and the environment. With the recognition that the watershed approach furnishes an effective strategy for the restoration of impaired ecosystems and protection of waters from pollution comes the need for tools to identify and quantify the contributions of different land uses to pollutant loading, as well as evaluating the effectiveness of various management practices in controlling different sources of pollution. To this end, the role of extant process-based hydrologic and water quality models in water resources planning and environmental management cannot be overemphasized. However, models are not immune from errors, and their predictions are subject to uncertainty arising from our limited conceptual understanding of complex natural systems and imperfect data.

Traditionally, the role of mathematical models, empirical or process-based, has been limited to testing scientific hypotheses, predictions, and interpreting experimental and field data. Recently, however, models are being used as tools for decision making and formulation of environmental policy. In addition to

[1]Conference speaker

forecasting and common applications, hydrologic and water quality models are also used in ecological sciences (e.g., Canham *et al.,* 2003; and Wu *et al.,* 2006) and environmental management and regulation, e.g., in risk assessment (USEPA, 2004), Regulatory Impact Analysis (RIA) (Krupnick *et al.,* 2006, Circular A-4, 2003), Total Maximum Daily Loading (TMDL) program (e.g., USEPA, 1999), and in many other enacted environmental laws. In almost every category, uncertainty analysis plays a critical role. Successful risk management and the formulation of environmental policy are hinged on realistic estimates of uncertainty (NRC, 2004; and Reckhow, 1994 (a), (b)). Failure to communicate model uncertainty, however, can lead to undesirable environmental consequences, with societal and economic implications (see Oreskes, 2003, citing the catastrophic 1997 Red River flooding in North Dakota).

The TMDL program, implemented at the watershed scale, is one of the most important environmental laws in the United States enacted to regulate point source and nonpoint sources of pollutant discharges to the nation's waters. Analysis uncertainty in the TMDL calculation is accounted for by applying a margin of safety (MOS). However, in currently practiced TMDLs, MOS is arbitrarily selected, and the degree of protection provided by the selected value is often unknown (NRC, 2001). Formal methods for the estimation of TMDLs and MOS on the basis of model prediction uncertainty are lacking. While the use of probabilistic techniques to characterize uncertainty has been promoted in risk assessment (e.g., USEPA, 2004), they have been limited to analyses *a posteriori* to model calibration. Further, rigorous applications of probabilistic methods to risk management have yet to find their way to environmental protection and regulation.

This paper identifies and discusses advantages and limitations of prominent approaches for uncertainty estimation in hydrologic and water quality modeling. It also presents a probabilistic framework that can be applied to risk management, in general, and TMDL calculation, in particular. The application of the risk management paradigm is illustrated by means of a hypothetical lake-sediment phosphorus TMDL calculation. A formal procedure is also presented for the estimation of the TMDL and associated MOS, as a function of model predictive uncertainty and desired degree of protection. The paper ends with conclusions and future challenges in research and applications of model uncertainty estimation to environmental decision making.

UNCERTAINTY ESTIMATION METHODS IN HYDROLOGIC MODELING

First-Order Approximation

The First-Order Approximation method (FOA) is a widely used methodology for estimating model output uncertainty due to its relative simplicity. However, related applications have been limited to post model calibration where model parameters are perturbed either based on experimental evidence or judgment. FOA can be used to estimate the mean and variance of the dependent variable Y, due to uncertainty of the dependent variables, X (Haan, 2002; and Ang and Tang, 2007). Consider Y as a given model output (e.g., runoff, streamflow, pollutant concentration, etc.) and $X = (X_1, X_2, \ldots, X_n)$ as a vector of n random variables denoting uncertain model input variables (e.g., precipitation intensity) and parameters (channel roughness coefficient, soil hydraulic conductivity, SCS dimensionless curve number, etc.). Mathematically, any of the outputs of hydrologic and water quality models can be described by the general functional relationship,

$$Y = g(X) \qquad \ldots (1)$$

where g denotes either a statistical or empirical relationship relating Y to X, or a process-based, mechanistic model. Expanding Y in Taylor series around the means of n random variables and truncating higher order terms yields the following (first-order) approximations to the mean and variance of Y,

$$\mu_Y = E(Y) \approx g(\mu_X), \ \mu_X = E(X) \qquad \ldots (2)$$

$$\sigma_Y^2 = Var(Y) \approx \nabla g(\mu_X)^T \Sigma_X \nabla g(\mu_X), \ \Sigma_X = E(X X^T) \qquad \ldots (3)$$

where E is the expectation operator; μ_X is the mean of the random vector X; ∇ is the gradient operator; Σ_X is the covariance matrix of X; and T is the matrix transpose operator. Note that in Eqns. (2) and (3), g and ∇g are evaluated at μ_X. Eqn. 3 can be described in a more useful form as,

$$CV_y \approx \sqrt{\Theta^T \Pi \, \Theta} \qquad \ldots (4)$$

in which,

$$\Theta = \begin{bmatrix} CN_1 \\ CN_2 \\ \vdots \\ CN_n \end{bmatrix}, \ \Pi = \begin{bmatrix} CV_1^2 & \rho_{1,2} CV_1 CV_2 \cdots \rho_{1,n} CV_1 CV_2 \\ \rho_{2,1} CV_2 CV_1 & CV_2^2 \cdots \rho_{2,n} CV_2 CV_n \\ \vdots & \\ \rho_{n,1} CV_n CV_1 & \cdots \quad CV_n^2 \end{bmatrix} \qquad \ldots (5)$$

where $CN_i = (\mu_{Xi}/\mu_Y)\partial g(\mu_{Xi})/\partial X_i$; CV_i is the coefficient of variation of X_i; and ρ_{ij} is the cross-correlation between X_i and X_j. Θ is a vector whose typical element CN_{Xi} is a measure of model sensitivity to X_i and is referred to as *condition number* (Chapra, 1997).

Besides computational efficiency and relative simplicity, one of the advantages of Eqn. (4) is its utility in estimating the relative contribution of each random variable to the overall spread (uncertainty) of the model output. An interesting point worth emphasizing is that Eqns. (4) and (5) transparently distinguish parameter sensitivity from parametric uncertainty, an issue that has had its share of confusion in the literature. The former is solely measured by CN_i and is essentially deterministic in nature, whereas the latter is measured by the product of CN_i and CV_i and thus inherently stochastic. Unless highly uncertain, the most sensitive parameter may not dominate model output uncertainty (e.g., Hantush and Kalin, 2005). Depending on the value of the *condition number*, CN_i, corresponding parameter uncertainty might either be magnified or attenuated during model computation of the output variance.

In spite of the widespread applications of the FOA in hydrologic modeling (e.g., Lee and Mays; 1986; Lei and Schilling, 1994; and Johnson and Rinaldi, 1998) and water quality modeling (e.g., Zhang and Shaw, 2004; Bobba *et al.*, 1996; Melching and Yoon, 1996; and Warwick, 1997), the errors committed by neglecting higher order terms in the Taylor series expansions can be significant for highly nonlinear systems, such as complex watershed models.

Generalized Likelihood Uncertainty Estimation

Recognizing model structural errors, parameter interactions, and nonlinearities inherent in many hydrologic models, GLUE method replaces the concept of the "optimum parameter set" with the concept of "equifinality" (Beven, 1993, and 2006). The latter is used in the sense that many different model structures and many different parameter combinations (behavior sets) within a chosen model structures are valid (acceptable) simulators of the observed system behavior. The methodology starts with the generation of an ensemble of parameter sets by sampling random parameter values from their respective prior distributions and conducting a simulation for each parameter set. A likelihood measure is then computed for each simulation and corresponding parameter combination by comparison of the results with observations using, e.g., Nash-Sutcliffe efficiency criterion (Beven and Binley, 1992),

$$L(\Theta_i|Y) = 1 - \frac{\sigma_i^2}{\sigma_o^2} \quad \ldots (6)$$

where $L(\Theta_i|Y)$ is a measure of the likelihood of simulating the data set **Y** given the parameter set Θ_i. σ_i^2 is the variance of residual errors; and σ_O^2 is the variance of observations. $L(\Theta_i|Y)$ is used here as a fuzzy measure of the probability of how well the parameter set allows the model to describe the data; a generalization in a sense to the *likelihood function* often used in classical Bayesian estimation. Once the likelihoods of all parameter sets are computed, those parameter sets whose likelihoods exceed or equal to a prescribed threshold value, 0 is typically chosen, are retained as *behavior* sets, and those which fail the test are assigned zero likelihood weights and disregarded as *nonbehavior* sets; i.e., the corresponding model output is dissimilar to the behavior of the system under study. The likelihood weights of the retained simulations are then rescaled so that the sum of their totals is 1.0. The scaled likelihood weights then are combined with the *priori* parameter distribution using Bayes theorem to yield the *posterior* likelihood weights,

$$\hat{f}(\Theta_i|Y) = c\,L(\Theta_i|Y)\,\hat{f}(\Theta_i) \quad \ldots (7)$$

where Θ_i is model parameter set i; $\hat{f}(\Theta_i|Y)$ is the *posterior* likelihood measure of Θ_i given the set of observations Y; $\hat{f}(\Theta_i)$ is the *prior* likelihood measure of Θ_i; and c is a normalization factor. The collective values of $\hat{f}(\theta_i|Y)$ over all parameter sets could be used to estimate the marginal probability density function (*pdf*) for model parameter θ.

The GLUE methodology constitutes a challenge to traditional approaches that embrace the concept of optimal parameter set in model calibration. Aside from acknowledging that models are at best only approximations of the real world and that model structural errors are implicitly accounted for in the methodology (Beven and Binely, 1992), the methodology provides a robust approach for the calibration and uncertainty estimation of highly nonlinear and complex environmental models, such as the case in watershed modeling. The nuisance of conventional gradient-based techniques in search of optimal solution in highly nonlinear, complex response surfaces that have several local optima is remedied by the global sampling nature in GLUE. One other advantage of the methodology is that by emphasizing parameter sets rather than individual parameters, the effect of interactions (covariation) among the parameters will

be reflected implicitly in the value of the likelihood measure associated with each set (Beven and Binely, 1992, and Schulz *et al.,* 1999). The methodology, however, is not without drawbacks, most notably is the subjectivity in selecting the likelihood measure and threshold criterion separating the *behavior* from *nonbehavior* parameter sets. Different values of the threshold can lead to different size of the output uncertainty band (Schulz *et al.,* 1999; and Zheng and Keller, 2007). The implication of arbitrarily selected likelihood measure is that GLUE's uncertainty limits are no longer direct estimates of the probability of observing a particular observation.

Classical Bayesian Estimation

Another widely applied Bayesian approach is the classical Bayesian analysis, henceforth, referred to as the Bayesian-error analysis (BEA). The BEA approach recasts a deterministic model into a standard regression form and conducts model simulations based on Bayesian statistics to estimate uncertainties (e.g., Sorooshian and Dracup, 1980; Bates and Cambell, 2001; Vrugt *et al.,* 2003a; Kavetski *et al.,* 2006; Ajami *et al.,* 2007; and Samanta *et al.,* 2007). With additive, typically Gaussian white noise error term, observed/ measured quantity Y_i can be expressed as,

$$Y_i = Y(X,\theta) + \varepsilon_i,\ \varepsilon_i \sim N(0,\sigma_\varepsilon^2),\ i = 1, 2,..., T \quad \ldots (8)$$

where $Y(X, \theta)$ is the model output; X denotes input variables; θ is a set of model parameters; ε_i is zero-mean error with constant but unknown variance, σ_ε^2; and T is length of observed record. In BEA, θ and σ_ε are treated as probabilistic variables having a joint posterior *pdf.* The likelihood function is derived with the assumption that the errors are independent identically distributed (*iid*), often but not always, Gaussian and homoscedastic (i.e., with a constant variance). For *iid* normal ε_i and uniform (non-informative) prior θ, using Bayes theorem and assuming a nonuniform prior of the form $p(\theta, \sigma_\varepsilon^2) \propto 1/\sigma_\varepsilon^2$, the joint posterior distribution of θ and σ_ε^2 may be expressed as follows (Samanta *et al.,* 2007),

$$P(\theta,\sigma_\varepsilon^2 \mid Y) \propto \sigma_\varepsilon^{-(n+2)} \prod_{i=1}^{n} \exp\left\{-\frac{1}{2\sigma_\varepsilon^2}[Y_i - Y(X,\theta)]^2\right\} \quad \ldots (9)$$

where n here is the observed data sample size. The error variance, σ_ε^2, can be estimated directly by applying the maximum likelihood method to the likelihood function (e.g., Thiemann *et al.,* 2001). Maximizing the likelihood function produces the output with the highest probability of being closest to the true value of the variable being forecasted. Thiemann *et al.* (2001) presented a Bayesian Recursive Estimation Algorithm (BaRE) through which various posterior and conditional densities are approximated via MC simulation and maximum likelihood estimation of the variance of measurement errors in transformed output space. The BaRE algorithm can be used to compute the probability density of the output measurement and update the forecast as new observation becomes available. Further, the approach could be used to construct posterior parameter set density and update for new measurements. Another approach is to sample posterior distribution in Eqn. 9 using the Markov Chain Monte Carlo (MCMC) simulation (e.g., Samanta *et al.,* 2007). In the MCMC method, prior (or proposal) parameter distributions are varied subject to conditions that ensure convergence to the posterior target distributions (Vrugt *et al.,* 2003a).

The main advantage of the BEA over GLUE is that in the former method marginal and conditional densities are manipulated according to the rules of mathematical probability embodied in Bayesian inferences without the ambiguity of admitting any "likelihood measure" $L(\Theta_i|Y)$ as an estimate to the "likelihood function" or $P(Y\ |\Theta_i)$, as the case in the GLUE methodology (Thiemann *et al.,* 2001). On the other hand, BEA aims at estimating the quantity $P(Y|\Theta_i)$ exactly, of course, after assuming *iid*, typically Gaussian residual errors. The possibility of nonconvergence of Markov chains to stationary distributions is a potential shortcoming in BEA method. On the other hand, GLUE is relatively easier to implement and computationally simpler.

Pareto Optimality

The Pareto Optimality method is inherently multi-objective in nature (Gupta *et al.,* 1998; Madsen, 2000; and Vrugt *et al.,* 2003b) and shares the parameter sets equivalence of the *equifinality* concept of the GLUE methodology in the sense that there are multiple ways in which the best fit of a model to observed data can be defined. The multiobjective equivalence of the parameter sets is more commonly referred to as *Pareto Optimal.* Similar to the *behavior* and *nonbehavior* sets in GLUE, the parameter space in Pareto Optimality can be partitioned into *good* or *Pareto* solutions and *bad* solutions. The number of the *good* parameter sets defines model output uncertainty.

In the most general form, the identification of the equivalent parameter sets may be cast in a multi-objective optimization problem,

$$\min_{\theta} F(\theta) = \{f_1(\theta), f_2(\theta), \ldots, f_m(\theta)\} \quad \ldots (10)$$

where F is a transformation of residual errors (e.g., min, weighted-sum of squares, Nash-Sutcliff coefficient, etc.) of one or multiple model outputs evaluated using the parameter values θ. The functions $f_i(\theta)$, i = 1, 2,..., m, are selected such that they are unrelated in the sense that they measure different model output residuals, or different important aspects of differences between the observed data and model simulations that can be used to extract the useful information contained in the data and transform it into estimates for the parameters (e.g., Gupta *et al.,* 1998). The solution to the multiobjective optimization problem (10) will consist of a Pareto parameter space Θ corresponding to various trade-offs among the objectives. Alternatively, the weighted sum of the objectives $f_i(\theta)$, i = 1, 2, ..., m can be minimized to obtain the Pareto parameter space Θ,

$$\min_{\theta} F(\theta) = \sum_{i=1}^{m} w_i f_i(\theta), \ \sum_{i=1}^{m} w_i = 1 \quad \ldots (11)$$

where w_i are weights whose values determined either randomly or in some other fashion generate the Pareto parameter sets by solving the essentially single-objective optimization problem (11) for each set of the weighting coefficients (e.g., Yan and Haan 1991). The shuffled complex evolution (SCE-UA) global optimization has been cited as effective algorithm for generating as many discrete Pareto solutions as necessary for (10) or (11) to obtain an acceptable approximation of the Pareto parameter space Θ (e.g., Duan *et al.,* 2007; and Kuczera, 1997).

The multiobjective property of the method may lead to identification of parameter values that are more general in the scope of applications and better describing the overall characteristic of a particular model output. The Pareto Optimality, however, is inherently a search-based technique as it requires the identification of the Pareto optimum. Even with the multiobjective complex evolution (MOCOM-UA) algorithm, which alleviated much of the computational burden associated with SCE-UA technique in the solution for the Pareto optimum (Θ), its search-based nature makes it more computationally demanding than the GLUE methodology. Further, the selection of F(θ) in (10) and w_is in (11) remains subjective in nature.

Kalman Filter

Although received relatively much less attention than the above described approaches in watershed and water quality modeling, Kalman filtering, especially Ensemble Kalman Filter (EnKF) (Evensen, 2003) has been extensively used in sequential data assimilation to, among others, atmospheric and climate forecasting, land-surface models, marine ecosystem models, and oceanographic problems. Although less applied to complex watershed and water quality models, EnKF approach holds greater promise in forecasting and conditional simulation in both areas. Kalman filtering is a predictor-corrector recursive algorithm that was originally developed for linear filtering problems (Kalman, 1960; and Jazwinski, 1970), and has been enhanced as the Extended Kalman Filter (EKF) for nonlinear systems (Lewis, 1986). After initializing the estimate of the state variable and associated covariance structure, Kalman filtering recursions start with a time update process for predicting state variable(s) and associated error covariance(s), followed by a measurement update (or correction) process for the assimilation of observed data into the system variable(s). Application of linear filters and EKF in hydrology has been limited to rainfall-runoff generation (e.g., Bras and Rodriguex-Iturbe, 1985), simple water quality models (e.g., Lettenmaier and Burges, 1976), and groundwater flow under conditions of uncertainty (Hantush and Mariño, 1994, 1997). Only recently, with the advent of EnKF (1992), applications of Kalman Filter to hydrology at the larger watershed scale gained momentum (e.g., Reichle *et al.,* 2002; Gabriëlle *et al.,* 2007; and Kim *et al.,* 2007). The EnKF nonlinearly propagates a finite ensemble of model trajectories, from which an estimate of the state(s) pdf(s) gives a complete statistical description of the state(s). These characteristics makes the EnKF easier to apply than EKF and more suitable for highly nonlinear problems.

In general, for higher dimension state variables, Kalman filters are computationally demanding and require accurate knowledge of first-two moments of the process and measurement noise. Common applications of Kalman filters are limited to white process and white measurement noise; autocorrelations in the process and measurement noise and cross correlation among the noise terms complicate the analysis. Its worth noting that despite the advantage of EnKF in the nonlinear projection of the state variable, the recursive relationship it utilizes during the measurement update remains an approximation derived from a linear state-space representation.

Other Methods

There are numerous methods worthy of reporting that overlap with and share some of the philosophy of the

major approaches described above. We were not able to report these to conserve space (e.g., Hantush and Kalin, 2005, 2008; Carpenter and Georgakakos, 2004; and van der Perk and Bierkens, 1997).

SCIENTIFIC UNCERTAINTY AND ENVIRONMENTAL REGULATION

In spite of reported successes in controlling point sources pollutant discharges in the US, water quality impairments continue to exist in the Nation's waters primarily due to diffuse nonpoint sources from agricultural runoff, forestry, land development activities, and urban runoff (USEPA, 1999). To address the combined, cumulative impacts of both point and nonpoint sources, the United States Environmental Protection Agency (USEPA) has adopted a watershed approach, of which total maximum daily loads (TMDLs) are a part. Section 303(d) of the Clean Water Act (CWA) requires states, territories, and authorized tribes to develop TMDLs for pollutants in impaired waters, including those which are threatened to be impaired also. A TMDL is the maximum of point and nonpoint source loads that can enter a water body without exceeding specified water quality standards.

The Comprehensive Environmental Response and Liability Act (CERCLA), otherwise known as Superfund, was enacted to address the legacy of contamination from chemical waste (Stephens, 2008). Numerous water quality regulations were put in place to achieve compliance and protect groundwater and surface water resources, most notably are the Safe Drinking Water Act, The Toxic Substance Control Act, and the Resource Conservation and Recovery Act (RCRA). Other enacted environmental laws include The Farm Security and Rural Investment Act which, among others, aims at protecting water quality from agricultural nonpoint source pollution. Of these environmental laws, the TMDL program is now considered to be pivotal in securing the nation's water quality goals (NRC, 2001); it establishes the allowable loadings, thereby providing the basis for states to establish water quality-based controls, the objective of which is attainment of ambient water quality standards through the control of both point and nonpoint sources of pollution. In almost all of these environmental laws, simulation models are used as to tools for regulatory compliance, risk assessment, and environmental planning and management.

In the following subsections, we establish the link between model predictive uncertainty and risk management and demonstrate its applicability to a hypothetical TMDL problem with a steady-sate lake-sediment phosphorus model.

Model Uncertainty and Regulatory Risk Management

Similar to the risk of failure of a given system capacity to sustain applied load, water quality compliance can probabilistically be described by this risk inequality,

$$\underbrace{P\{Y \geq Z\}}_{\substack{\text{Risk}\\ \text{Assessment}}} \quad \underbrace{\leq \beta}_{\substack{\text{Risk}\\ \text{Management}}} \qquad \ldots (12)$$

where Y is pollutant concentration [ML^{-3}] or loading rate [MT^{-1}] measured or computed by an empirical relationship or a process-based mathematical model; Z is the compliance concentration [ML^{-3}] or loading rate [MT^{-1}], such as the Maximum Contaminant Level (MCL) and TMDL designated use criterion; and β is the acceptable level of risk or frequency of violations ($(1-\beta)$ being the confidence level or degree of protection). According to the U.S. EPA (1997), for the waterbody to be listed as unimpaired, no more than 10% of the samples ($\beta = 0.1$) collected from the waterbody should violate water quality standards. In general, Z also is subject to uncertainty, although, fixed values are common in practice. For example, in TMDLs, the uncertainty in Z depends on the criterion selection process and its accuracy as a surrogate measure to the designated use of the waterbody (NRC, 2001). Y is random due to measurement and modeling errors combined.

The above inequality articulates the risk-assessment-risk-management paradigm in tractable mathematical form. The first term on the left-hand-side of (12) (i.e., $P\{Y \geq Z\}$) concerns the estimation of the risk which is primarily a science issue, whereas the right-hand-side of the inequality constitutes the risk management component. The selection of acceptable risk value (β) is a policy matter that is left to decision makers. Eventually, risk assessment is integrated with other considerations in order to make and justify regulatory decisions (U.S. EPA, 2004). Federal, local States, tribes, and academic institutes in the USA together are involved in efforts aiming at criteria development and the selection of water quality numerical targets (U.S. EPA, 2000).

Assuming that Z is fuzzy, varies in the range $[a, b]$, and statistically independent of Y, it can be shown that,

$$\begin{aligned} P\{Y > Z\} &= 1 - \int_a^b \int_0^z f_{Y,Z}(y,z)\,dy\,dz \\ &= 1 - \int_a^b F_Y(z)\,f_Z(z)\,dz \end{aligned} \qquad \ldots (13)$$

where $F_Y(z)$ is the cumulative distribution function of Y; and $f_Z(z)$ is probability density function of Z. Equation (13) states that for a fuzzy compliance criterion, the probability of compliance is equal to weighted average of the cumulative probability distribution of measured/computed water quality indicator evaluated over all possible values of Z, with the weights reflecting our degree of belief in rather a fuzzy numerical water quality target.

To illustrate the applicability of Eqn. (12) to a TMDL calculation, we consider a simple lake-sediment phosphorus model and limit the analysis to brute force application of MC method.

Application to Lake Phosphorus TMDL

Figure 1 depicts the conceptual lake-sediment phosphorus model. The lake is divided into a water layer and bottom sediments. Flow within the lake is assumed steady with no net loss by evaporation and groundwater seepage. Sediments are transported into and out of the lake by advection and are subject to deposition to and resuspension from bottom sediments. The bottom sediment layer receives depositional flux, releases some of the sediments back to the water column, and is buried by sedimentation. Phosphorus is usually highly sorbed to sediment particles and the dissolved component is assumed to be at equilibrium with the sorbed phase in both the lake water and the bottom sediment. Mass transfer of dissolved phosphorus is assumed to occur by diffusion across the sediment-water interface.

Mathematically, the lake sediment-phosphorus hypothesis can be posed in the following ordinary differential equations for the lake and sediment environments:

Mass balance of sediment in lake,

$$V\frac{dm_w}{dt} = Q_{in}\,m_{w,in} - v_s A_s\,m_w + v_r\,A_s\,m_s - Q_{out}\,m_w \quad \ldots (14)$$

Mass balance of sediment in lake-bottom,

$$A_s\,H\frac{dm_s}{dt} = v_s\,A_s\,m_w - v_r\,A_s\,m_s - v_b\,A_s\,m_s \quad \ldots (15)$$

Mass balance of total phosphorus in lake water,

$$V\frac{dC}{dt} = Q_{in}\,C_{in} - v_s\,A_s F_s C + \alpha\,A_s(f_d P - F_d C) + v_r A_s\,f_s P - Q_{out}\,C \quad \ldots (16)$$

Mass balance of dissolved phosphorus in sediment,

$$V_s\frac{dP}{dt} = v_s\,A_s F_s C - \alpha\,A_s(f_d P - F_d C) - v_r A_s f_s P - v_b A_s P \quad \ldots (17)$$

where m_w is sediment concentration in lake water [ML^{-3}]; $m_{w,in}$ is inflow sediment concentration [ML^{-3}]; m_s is lake-bottom sediment concentration [ML^{-3}]; C is total phosphorus concentration in lake water [ML^{-3}]; P is total phosphorus concentration in sediment layer [ML^{-3}]; V is lake water volume [L^3]; Q_{in} is lake inflow rate [L^3T^{-1}]; Q_{out} is lake outflow rate [L^3T^{-1}]; A_s is lake surface area at the sediment-water interface [L^2]; H is sediment layer thickness [L]; V_s is volume of sediment layer [L^{-3}]; v_s is settling velocity [LT^{-1}]; v_r is resuspension/recycling velocity [LT^{-1}]; v_b is burial velocity [LT^{-1}]; and α is effective mass-transfer velocity coefficient at the sediment-water interface [T^{-1}]. For illustrative purposes, we assume $\alpha = D^*/\delta$, where D^* is free-water phosphorus diffusion coefficient [L^2T^{-1}]; and δ is thickness of sediment-water interface diffusion boundary layer [L].

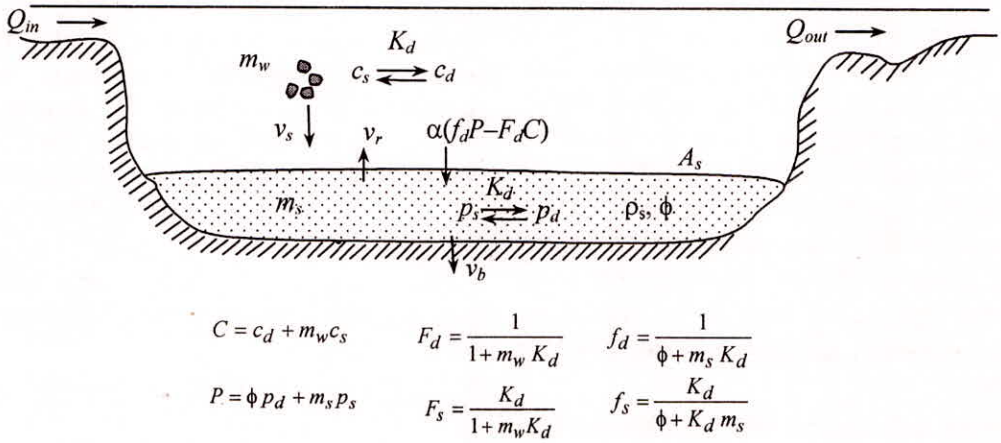

Fig. 1: Conceptual lake-sediment phosphorus model

Assuming $Q_{in} = Q_{out} = Q$ and taking all time-derivatives equal to zero, the steady-state solution of the sediment equations can be shown to be,

$$m_w = m_{w,in} - \frac{v_b A_s}{Q}(1-\phi)\rho_s \quad \ldots (18)$$

For sediment concentration in the lake, and for total phosphorus, the solution is,

$$C = \frac{\alpha f_d + v_r f_s + v_b}{v_b A_s (v_s F_s + \alpha F_d) + Q(\alpha f_d + v_r f_s + v_b)} W \quad \ldots (19)$$

in which,

$$C = c_d + m_w c_s \quad \ldots (20a)$$

$$W = Q c_{d,in} + Q m_{w,in} C_{s,in} \quad \ldots (20b)$$

where W is total phosphorus loading rate [MT^{-1}]; ϕ is sediment layer porosity [L^3L^{-3}]; ρ_s is sediment particle density [ML^{-3}]; c_d is dissolved phosphorus concentration in lake water [ML^{-3}]; c_s is sediment-bound phosphorus concentration in lake water [MM^{-1}]; $c_{d,in}$ is inflow dissolved phosphorus concentration [ML^{-3}]; $c_{s,in}$ is inflow sediment-bound phosphorus concentration [MM^{-1}]; and f_d, f_s [L^3M^{-1}], F_d, and F_s [L^3M^{-1}] are partition coefficients relating dissolved and sorbed-phase phosphorus concentration to total phosphor concentration (equations shown in Figure 1, where p_d is pore-water phosphorus concentration in bottom sediments [ML^{-3}]; p_s is sorbed phosphorus concentration in bottom sediments [MM^{-1}]; and K_d is phosphorus sorption coefficient [L^3M^{-1}]).

A more direct measure to eutrophication, which reflects the nuisance conditions caused by excessive algal blooms, is chlorophyll a, *Chl a*. Biocriteria, such as *Chl a*, are better indicators of designated uses than are chemical criteria and their use in TMDL development should be promoted (NRC, 2001). Without loss of generality, we select *Chl a*-phosphorus regressed relationship reported by Schnoor (1996) (see Chapra, 1997 for other relationships),

$$Y = 0.081\, C^{1.46}\,.\, R = 0.95 \quad \ldots (21)$$

where Y denotes *Chl a* concentration; both Y and C have the units of μg/L (mg/m^3).

Table 1 lists the parameters of the lake-sediment phosphorus model (Eqns. 19–20), their units, and assumed probability distributions and related parameters selected from typical values and ranges reported in the literature. A hypothetical bottom lake surface area, A_s, of 5×10^6 m^2 is assumed. Dissolved phosphorus is assumed to be 5% of the sorbed phosphorus in (g/m^3): $c_{d,in} = 0.05\, m_{w,in}\, c_{s,in}$. Z is *Chl a* enforced water quality target concentration; it is assumed to be a random variable and follows a triangular distribution with minimum, most probable (mode), and maximum values of 2.7, 10, and 14 μg/L, respectively. $P\{Y > Z\} > 0.1$ implies a eutrophic condition. If $P\{Y > Z\}$ is greater than 0.1, then for the TMDL to be achieved with 90% compliance, total phosphorus loading rate, W, should be reduced such that $P\{Y > Z\}$ is at least equal to 0.1.

Vollenwieder and Kerekes (1980) reported *Chl a* concentration between 2.7 and 78 μg/L as the range for eutrophication and 3 to 11 μg/L for mesotrophic status. The overlap in Vollenwieder and Kerekes (1980) trophic status classification scheme boundaries, the variability of selected target values among states, and variation in the methodology for selecting a numerical target altogether provide a rationale for treating water quality numerical targets (i.e., compliance values) as fuzzy variables whose probabilities express the degree in one's belief that compliance based on particular Z values would restore or prevent the degradation of a targeted water body.

Table 1: Lake Sediment-Phosphor Model Parameters and Assumed Probability Distributions

Parameter	*Unit*	*Distribution*	*Parameter*	*Unit*	*Distribution*
Q	m^3/yr	~N(2×10^8, 2×10^6)[a]	D^*	m^2/yr	~U(1.5, 2.32)[d]
$m_{w,in}$	g/m^3	~N(35, 5)	δ	mm	~log-N(1.0, 1.0)
$c_{s,in}$	g/g	~log-N(10^{-3}, 4×10^{-4})[b]	ϕ	m^3m^{-3}	~T(0.4, 0.7, 0.95)
v_s	m/yr	~log-N(37, 20)	ρ_s	g/cm^3	~T(1, 2.6, 2.7)
K_d	m^3/g	~log-N(4.87×10^{-4}, 2.59×10^{-5})	Z	mg/m^3	~T(2.7, 10, 14)
v_b	m/yr	~T(8×10^{-4}, 0.005, 0.01)[c]			

[a] N(λ, σ) denotes normally-distributed random variable with mean λ and standard deviation σ.
[b] log-N(λ, σ) denoted log-normally distributed random variable with mean λ and standard deviation σ.
[c] T(a, b, c) denotes triangular distribution of random variable with minimum value a, mode c, and maximum value c.
[d] U(a, b) denotes a uniformly distributed random variable in the interval (a, b).

For each MC simulation 100,000 independent parameter sets were sampled by randomly generating parameter values from their respective pdfs in Table 1. An ensemble of model outputs was generated with 100,000 model runs of Eqns. (19–21), and an ensemble of same size was generated for Z from its pdf in Table 1. The MC simulation and the sampled Z yielded $P\{Y \geq Z\} = 0.73$. A reduction in total phosphorus loading rate, W, was therefore needed to reduce $P\{Y \geq Z\}$ to 0.1 and achieve the TMDL at 90% compliance. From Eqn. (20b); it is evident that either or both of the two variables $m_{w,in}$ and $c_{s,in}$ will need to be manipulated to achieve this target (recall, $c_{d,in}$ is assumed to be a fixed fraction of $c_{s,in}$).

Figure 2 depicts the results of two selected TMDL strategies; the first seeks the reduction of sediment loading concentration, $m_{w,in}$, and the second involves the reduction of sediment-bound phosphorus concentration, $c_{s,in}$. MC simulation and a nonlinear-based root's finding technique were implemented to identify the mean $m_{w,in}$, or mean $c_{s,in}$ such that $P\{Y \geq Z^*\} = 0.1$. The variance of both management-control variables was held constant during the search procedure. The search procedure was programmed and executed in MATLAB computational package.

The upper panel (a) in Figure 2 shows relative frequency (100 bins) and cumulative distribution functions of precompliance sediment loading, $m_{w,in}$, and those required to achieve the phosphorus TMDL at 90% confidence (i.e., $\beta = 10\%$). The lower panel (b) shows similar output for the input sediment-bound phosphorus concentration, $c_{s,in}$. For the first strategy, an estimated 62% reduction in mean total phosphorus loading rate, W, was needed to meet the 10% requirement; whereas the second strategy yielded a higher 69% reduction in W.

The increased required load reduction in the second strategy compared to the first highlights the importance of parametric uncertainty on model estimated TMDL. Inspection of data in Table 1 shows higher uncertainty in $c_{s,in}$ than $m_{w,in}$: the coefficient of variation associated with the former is almost three times higher than the latter. Although not shown in a figure, the relative frequency of the computed phosphorus TMDL was relatively more skewed to the left for the second strategy, in favor of higher probability densities for lower phosphorus loading rate values, than for the first strategy; again, this is attributed to higher uncertainty in $c_{s,in}$.

A robust, perhaps more economically and socio-politically acceptable, approach might have involved the simultaneous reduction of both $c_{s,in}$ than $m_{w,in}$. Various management strategies could be implemented to achieve the hypothesized phosphorus TMDL. On one end of the spectrum, phosphorus input from anthropogenic point and nonpoint sources could be managed such that their discharges to watersheds and receiving water bodies are reduced. Example strategies

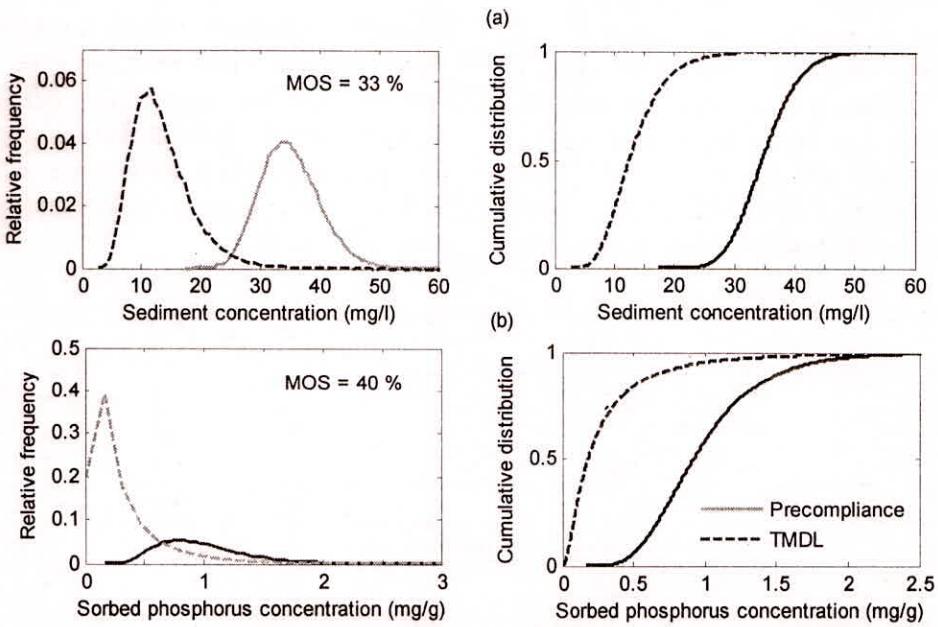

Fig. 2: Monte Carlo simulated precompliance (solid-line) and 10% compliance ($\beta = 0.1$) (dashed-line) relative frequencies and cumulative distributions: (a) Loading sediment concentration, $m_{w,in}$; and (b) loading sorbed phosphorus concentration, $c_{s,in}$. Results are based on MC simulation with 100,000 model runs

include improved agricultural practices such as increased nutrient retention in soils and crop uptake; limiting excess nutrients in fertilizer applications; and managing animal waste discharge to rangelands and agricultural fields. Enforcing the National Pollutant Discharge Elimination System (NPDES) insures nutrients point source load reductions. On the other side of the spectrum, sediment loading could be reduced by implementing sound management practices that reduce both soil and channel erosion and trap sediments that are major carriers for sorbed-phase phosphorus. Management strategies that combine both ends, i.e., controlling both sediment and phosphorus inputs with relatively varying degrees, may prove to be the most efficient and effective.

TMDL MARGIN OF SAFETY (MOS)

Traditionally, MOS is used to account for uncertainty in the relationship between pollutant loads and receiving water quality. Explicit and implicit approaches have been reported to have been used for the estimation or selection of MOS (Dilks and Freedman, 2004; and Shirmohammadi *et al.,* 2006). The most common approach for the estimation of MOS is the explicit selection of MOS as a percentage of the TMDL, a concept that is equivalent to a safety factor. The implicit approach is another widely used approach in which the margin of safety is introduced implicitly through the use of conservative assumptions in calculating the allowable load. Only a few studies considered uncertainty analysis as a base for estimating MOS (e.g., Borsuk *et al.,* 2002; and Zhang and Yu, 2004).

Its worth noting, however, that in the course of calculating the TMDL for the above hypothetical lake-sediment phosphorous model, formal account of model uncertainty was made (assuming parameters are the only source of uncertainty), but without having to make an explicit reference to the margin of safety (MOS). For completeness, however, a MOS may be computed as a function of the desired degree of protection, but in a more rigorous and formal way than the currently practiced implicit and explicit approaches.

The following is one way of calculating MOS in a relative sensem,

$$MOS = \frac{LR2 - LR1}{L_p} \times 100\,,\ LR1 = L_p - L_c,\ LR2 = L_p^* - L_c^* \qquad \ldots (22)$$

where L_p is precompliance load computed with expected parameter values (i.e., deterministic); L_c is compliance load computed with expected parameter values and expected Z; L_p^* is mean precomplinace load based on random parameters and Z; and L_c^* is mean compliance load based on random parameters and Z. Noting that L_p is a first-order approximation of L_p^*, and in this particular example, $L_p \approx L_p^*$, we have,

$$MOS \approx \frac{L_c - L_c^*}{L_p} \qquad \ldots (23)$$

Note that the reduction in the required mean phosphorus load is expressed relative to the deterministic precompliance load. For the data in Table 1 and MC simulation, MOS = 33% for the TMDL strategy involving $m_{w,in}$ as the management-control variable, whereas the strategy based on $c_{s,in}$ control yielded MOS = 40%; these results are consistent with the estimated larger mean load reduction required for the second strategy, as discussed above.

The variation of MOS with β is shown in Figure 3. The larger the value of β, the smaller the level of protection against violating the designated use numerical target, the smaller the MOS. Smaller β is associated with larger MOS, consequently, larger load reduction and costlier management actions. Assuming that MOS values can be translated into cost, the computed relationship in Figure 3 approaches the TMDL problem with an attitude toward risk and provides decision makers with the option of weighing risk of violating compliance targets versus the cost of impending management actions. Current practices in TMDL development and approaches for estimation of the MOS are inadequate at producing such an explicit relationship.

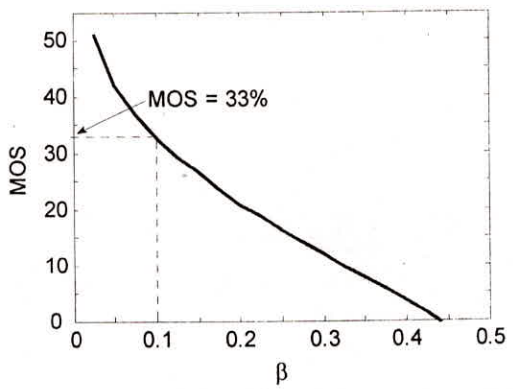

Fig. 3: MOS as a function of β. 1–β is the probability of compliance (confidence level)

It's worth noting that as new observations of *Chl a* and/or *P* are collected, a Bayesian update of the

TMDL distribution and associated MOS could be obtained. In this regard GLUE, BEA, and EnKF methodologies are ideally suited to address a TMDL within the probabilistic framework described above, since these methods produce an estimate of the probability distribution of the model output, *Y*. The value of data collection in improving a TMDL calculation through Bayesian deductions has not been addressed here or in the literature hitherto, but certainly constitutes an endeavor worthy of future research.

SUMMARY AND FUTURE DIRECTIONS

The use of simulation models as a tool for scientific analysis and environmental management and regulation is constrained by the inability of these models to replicate the exact systems' responses they aim to simulate. This problem is further confounded by uncertainty in the data available and lack thereof to calibrate these models in complex environmental systems. Research efforts, therefore, should be directed toward quantifying predictive uncertainty of models, and communicating that uncertainty to decision makers to evaluate the consequences of alternative actions and possible events.

Multitude of methods has been reported in the literature dealing with model uncertainty estimation. From these methods, we identified five prominent and promising approaches in hydrologic and water quality modeling, namely, First-Order Approximation (FOA), classical Bayesian Estimation (BEA), Generalized Likelihood Uncertainty Estimation (GLUE), Pareto Optimality, and the Ensemble Kalman Filter (EnKF). Merits and demerits of each approach were discussed. Among these approaches, the BEA, GLUE, and EnKF methodologies standout, and hold great promise for forecasting and probabilistic risk management in complex, highly nonlinear environmental systems. The multiobjective idea in Pareto Optimality is novel and can be extended to GLUE methodology. Both BEA and EnKF produce estimates of the *pdfs* of model outputs, whereas in FOA, GLUE, and Pareto Optimality, *pdf*s are crudely estimated. The disadvantage of assuming independence across parameter distributions often cited in traditional MC simulation can be remedied through Bayesian inferences in both BEA and GLUE, whereby prior distributions are updated by conditioning on observed data to produce posterior parameter sets that reflect covariations among the parameters.

It is known that models are unrealistic in modeling current conditions and are expected to produce poor predictions in ungauged watersheds, or when observations and measurements of input variables are sparse. However, it might also be argued that under such conditions models could still be used to predict the relative magnitude of change for different scenarios. To conserve space, we were not able to report lessons learned from past studies that support such an argument. Such case studies included the application of semidistributed watershed models to two different watersheds, one that is impaired by nutrients (Arabi *et al.*, 2007), and the other threatened by urbanization (Kalin and Hantush, 2006). These studies showed that the impact of Best Management Practices (BMPs) and land-use changes could be forecasted (with much less uncertainty) even when the models are relatively less accurate in modeling current conditions. The implication of this finding is that in ungauged watersheds, or when available measurements and input data are too sparse to calibrate models adequately, models could still be used to predict the relative magnitude of change for different scenarios. This argument may have implications on hydrologic applications dealing with long-term impacts of climate changes on watersheds' responses.

It appears that the literature is replete with research developments in uncertainty analysis, but they have been limited to simulation and forecasting, and they are yet to be fully explored in environmental risk management. The challenge now for science is to motivate decisions about appropriate management actions under conditions of uncertainty, as well as to aim at reducing the uncertainty by implementing effective monitoring programs. The TMDL program, mandated by section 303(d) of the CWA, requires uncertainty in the analysis be accounted for in a MOS. However, arbitrary selection of MOS appears to be the rule rather than the exception in current TMDLs. The proposed probabilistic framework, articulated through a hypothetical lake-sediment phosphorus TMDL example, provides an opportunity to enhance the risk assessment-risk management paradigm, in general, and provides an impetus to advance the TMDL program with more explicit account of model uncertainty and more formal estimation of the MOS. It was shown that MOS was independent of the probabilistic analysis, however, for completeness, we provided a formal, probabilistically-based approach for the estimation of MOS, as a function of modeling uncertainty, desired degree of protection (i.e., compliance probability), and fuzzy compliance target value.

It is anticipated that the demand for computer simulations in forecasting and environmental management will continue to grow; with this comes the importance of quantifying model predictive uncertainty. Robust uncertainty estimation methods are available, and the computing technologies to execute those methodologies and speedup their convergence are becoming more powerful and faster. In spite of the scientific and technological advances, much of the research in uncertainty estimation has been limited to hypotheses testing and scientific analysis, and very little has found its way to regulatory decision making. More research and development need to be accomplished to rigorously integrate model uncertainty into environmental decision making and formally quantify its importance in environmental risk assessment (e.g., superfund sites, landfills, nuclear waste management, etc.) and watershed management (e.g., TMDL programs). With this paradigm shift toward risk management, comes the challenge of improving scientific predictions and their relevance to societal values. Future challenges include, but not limited to, the following:

1. Uncertainty in hydrologic models stems from errors in model structure, parametric uncertainty, and input and measurement errors. Currently, we are unable to identify the relative importance of each source of uncertainty and propagate them in an explicit manner. A proper framework is lacking.
2. The predominant extant approach is to formulate a deterministic model, then employ an expensive Monte Carlo procedure to evaluate uncertainty effects. Are there other ways to assess and quantify uncertainty, i.e. can we have uncertainty information decide on the model? This may be possible in statistical models, but is still an area of active research.
3. With advances in sensor technology and use of satellite and other remote sensing instruments, we deal with massive amounts of data which comes with uncertainty. All this information cannot be utilized directly in models, and data compression and data mining become important issues. How are these methods to be employed if data have uncertainty? Standard methods require the data samples to be uncorrelated if not independent. Extensive spatial and temporal fields tend to be correlated both in time and space? What to do in such situations?
4. Often, we are interested in data that vary in time in a non-stationary fashion. Climate change effects are known to cause changes in rainfall and streamflow patterns. Meanwhile, we have now been measuring hydrologic data more frequently in space and time and with better instrumentation so that there exists heterogeneity not only in the data, but also in the uncertainty associated with the data. We need to develop smarter methods to tackle these problems.
5. Models are typically calibrated or the GLUE and BEA methodologies are performed using the observed data at the watershed outlet. The same models with their calibrated parameters or their parameter uncertainty ranges reduced are used for predictions in inner locations. The additional uncertainty introduced during this transition has not been formally addressed. There needs to be additional research in this area.
6. Uncertainties in land use/cover (LULC) change impact studies: In general a model is calibrated using the current LULC and observed data and then used in a predictive mode to study the effect of future projected LULC. Future LULC may have some LULC types not present in the past or the current LULC (consider transition from total forest to urban/forest). The uncertainty due to such set up remains to be resolved.
7. Interdisciplinary studies are becoming increasingly popular. National Science Foundation (NSF) has a program called Dynamics of Coupled Natural and Human Systems (CNH) which "promotes quantitative, interdisciplinary analyses of relevant human and natural system processes and complex interactions among human and natural systems at diverse scales" (quoted directly from NSF). Such studies require translation of data from one to another. Output of one discipline becomes input to another. As each discipline has different perspective of looking at the problems, there is an inevitable addition of uncertainty during the flow of information, which is commonly referred to as "transboundary uncertainty". Research in this field is still premature. Methods are needed for quantifying decision and linguistic uncertainty associated with communicating model results and uncertainty thereof to stakeholders and decision makers.
8. Uncertainty can never be eliminated due to the inherent spatiotemporal variability of environmental systems and data limitations, but it can be reduced. Research into methods reducing predictive uncertainty is needed: When, where, at what scale observations/measurements should be made?

Could uncertainty be reduced using multiple objectives or, perhaps, multiple models?

9. Integration of data and models uncertainty with societal values and economic benefits. One way to address this need is to integrate prediction uncertainty with a utility function to allow decision makers to maximize expected utility or maximum net benefits.
10. Current practices in regulatory risk assessment conduct uncertainty analysis a posteriori to model calibration and validation. Future approaches should focus on the integration of uncertainty estimation with model calibration; decoupling the two, and without the benefit of Bayesian update, could result in subjective and overly estimated uncertainty limits. Aside from sensitivity analysis, protocols need to be developed for the integration of uncertainty analysis into model studies of regulatory nature.
11. Estimation of TMDL uncertainty and MOS through formal, rigorous uncertainty analysis. Methods relating TMDL calculation and MOS to the desired degree of protection and uncertain compliance water quality targets are needed. Bayesian update of TMDL and MOS conditioned on new observations/measurements is worthy of further research.
12. Since biocriteria are a better indicator of designated use than are chemical criteria, future research challenge is to develop models, both mechanistic and empirical, that can more effectively link environmental stressors and control actions to biological response. The primary objective of this line of research is to reduce uncertainty in the TMDL calculation and required control actions.
13. Methods are lacking for estimating uncertainty in the performance of BMPs and impact of land use/land cover changes when data are not available.
14. Much of the literature in model uncertainty estimation have focused on parametric and model structural errors. Research is needed into methods resolving observation error from other sources of errors, incompatibility between observation and predicted variable scales, errors in boundary condition specification, and land-based and meteorological inputs. Effect of spatial heterogeneity and its relationship to scaling and effective parameters deserves further research.
15. Resolving the subjective choice of the likelihood measure and threshold criterion separating behavioural from nonbehavioural parameter sets in the GLUE methodology is still an active research area. Efficient sampling in the model parameter space and criteria for optimal sampling size for GLUE, BEA, and EnKF is another research area. While MCMC methods have removed the computational hurdle in BEA, its convergence to stationary distribution poses a challenge.
16. With the foreseen impact of global climate changes on water availability and quality, long-term model simulations will demand the coupling of global ocean-atmosphere circulation models with watershed models. With the highly complex global circulation models, application of the computationally demanding MC-based uncertainty estimation methods to the coupled watershed-atmosphere models is a formidable task deserving future research.
17. Research into improving processes representation in watershed models will continue, but, in hindsight, should only be within the scope of reducing uncertainty. Overemphasizing model complexity leads to increased number of parameters to be calibrated and increased requirement for observed data. Increasing physical realism does not necessarily translate into better model performance, especially if added parameters cannot be estimated independently. Reconciling improved model performance with the parameter identifiability problem as models increase in complexity is an issue deserving more research.
18. Research into outreach and education on the importance of knowledge about uncertainty is a first step for decision makers and policy formulators to demand it. Part of the education should be devoted to lessons learned from past failures to account for uncertainty during catastrophic events. Research into protocols promoting dialogues between modelers and managers may help reduce cost of analysis and transboundary uncertainty.
19. Much of the research in uncertainty analysis has been limited to forecasting and rainfall- runoff modeling. Insufficient research has been carried in TMDLs development and in the area of environmental management where optimal decisions regarding societal problems are sought.

ACKNOWLEDGEMENTS

The U.S. Environmental Protection Agency through its Office of Research and Development funded and managed the research described here through in-house efforts at the National Risk Management Research Laboratory. It has not been subjected to Agency review and therefore does not necessarily reflect the views of the Agency, and no official endorsement should be inferred.

REFERENCES

Ajami, N.K., Duan, Q. and Sorooshian, S. (2007). "An integrated hydrologic Bayesian multimodel combination framework: Confronting input, parameter and model structural uncertainty in hydrologic prediction". *Water Resour. Res.*, 43, doi:10.1029/2005WR004745.

Arabi, M., Govindaraju, R.S. and Hantush, M.M. (2007). "A probabilistic approach for analysis of uncertainty in evaluation of watershed management practices". *J. Hydrol.*, 333, 459–471.

Ang, A.H.-S. and Tang, W.H. (2007). *Probability concepts in engineering, Emphasis on applications in civil & environmental engineering*, John Wiley & Sons, Inc, NJ.

Bates, B.C. and Campbell, E.P. (2007). "A Markov chain Monte Carlo scheme for parameter estimation and inference in conceptual rainfall-runoff modeling". *Water Resour. Res.*, 37(4), 937–947.

Beven, K. (2006). "A manifesto for the equifinality thesis". *J. Hydrol.*, 320, 18–36.

Beven, K. (1993). "Prophecy, reality and uncertainty in distributed hydrologic modeling". *Adv. Water Resour.*, 16: 41–51.

Beven, K. and Binley, A. (1992). "The future of distributed models: Model calibration and uncertainty prediction". *Hydrol. Process.*, 6: 279–298.

Bobba, A.G., Singh, V.P. and Bengtsoon, L. (1996). "Application of first-order and Monte Carlo analysis in watershed water quality models". *Water Resour. Manage.*, 10, 219–240.

Borsuk, M.E., Stow, C.A. and Reckhow, K.H. (2002). "Predicting the frequency of water quality standard violations: A probabilistic approach for TMDL development". *Environ. Sci. & Technol.*, 36(10), 2109–2115.

Bras, R.L. and Rodriguez-Iturbe, I. (1985). *Random functions and hydrology*, Addison-Wesley, reading, Mass.

Carpenter, T.M. and Georgakakos, K.P. (2004). "Impacts of parametric and radar rainfall uncertainty on the ensemble streamflow simulations of a distributed hydrologic model". *J. Hydrol.*, 298: 202–221.

Canham, C.D., Cole, J.J. and Lauenroth, W.K. (2003). *Models in ecosystems science*, Princeton University Press, NJ.

Chapra, S.C. (1997). *Surface water-quality modeling*, McGraw-Hill, NY.

De Lannoy, G.J.M., Houser, P.R., Pauwels, V.R.N. and Verhoest, N.E.C. (2007). "State and bias estimation for soil moisture profiles by an ensemble Kalman filter: Effect of assimilation depth and frequency". *Water Resour. Res.*, 43, W06401, doi:10.1029/2006WR005100, 1–15.

Dilks, D.W. and Freedman, P.L. (2004). "Improved consideration of the margin of safety in Total Maximum Daily Load development". *J. Environ. Engrng.*, 130 (6), 690–694.

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

Evensen, G. (2003). "The ensemble Kalman filter: Theoretical formulation and practical implementation". *Ocean. Dyn*, 53, 343–367.

Gupta, H.V., Sorooshian, S. and Yapo, P.O. (1998). "Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information". *Water Resour. Res.*, 34(4): 751–763.

Haan, C.T. (2002). *Statistical methods in hydrology*, Second Edition, Iowa State Press, A Blackwell Publishing Company.

Hantush, M.M. and Kalin, L. (2008). "Stochastic-residual error analysis for estimating hydrologic model predictive uncertainty". *J. Hydrol. Engrng.*, 13(7), 585–596.

Hantush, M.M. and Kalin, L. (2005). "Uncertainty and sensitivity analysis of runoff and sediment yield in a small agricultural watershed with KINEROS2". *Hydrol. Sci. J.*, 50(6): 1151–1171.

Hantush, M.M. and Mariño, M.A. (1997). "Stochastic solution to Inverse problem in ground water". *J. Hydraul. Engrng.*, 123(12), 1139–1146.

Hantush, M.M. and Mariño, M.A. (1994). "Two-dimensional stochastic analysis and optimal estimation in aquifers: Random recharge". *Water Resour. Res.*, 30(2), 559–569.

Jazwinski, A.H. (1970). *Stochastic processes and filtering theory*, Academic, San Diego, California.

Johnson, P.A. and Rinaldi, M. (1998). "Uncertainty in stream channel restoration". *Uncertainty modeling and analysis in civil engineering*, Ayyub, B. M., ed., CRC Press, Boca Raton, FL., 425–437.

Kalman, R.E. (1960). "A new approach to linear filtering and prediction problems". *J. Basic. Eng.*, 82, 35–45.

Kavetski, D., Kuczera, G. and Franks, S.W. (2006). "Bayesian analysis of input uncertainty in hydrologic modeling: 1. Theory". *Water Resour. Res.*, 42, doi:10.1029/2005WR004368.

Kim, S., Tachikawa, Y. and Takara, K. (2007). "Applying a recursive update algorithm to a distributed hydrologic model". *J. Hydrol. Engrng.*, 12(3), 336–344.

Krupnick, A., Morgenstern, R., Batz, M., Nelson, P., Burtraw, D., Shih, J.-S. and McWilliams, M. (2006). "Not a sure thing: Making regulatory choices under uncertainty, Resources for the future", *Rep.*, Washington, D.C.

Kuczera, G. (1997). "Efficient subspace probabilistic parameter optimization for catchment models". *Water Resour. Res.*, 33, 177–185.

Lee, H.L. and Mays, L.W. (1986). "Hydraulic uncertainties in flood levee capacity". *J. Hydraul. Engrng.*, 112(10), 928–934.

Lei, J. and Schilling, W. (1994). "Parameter uncertainty propagation analysis for urban rainfall runoff modeling". *Water Sci. Technol.*, 29(1–2), 145–154.

Lettenmaier, D.P. and Burges. S.J. (1976). "Use of state estimation techniques in water resources system modeling". *Water Resour. Bull.*, 12(1), 83–99.

Lewis, F.L. (1986). *Optimal estimation with an Introduction to stochastic control theory*, John Wiley & Sons, NY.

Madsen, H. (2000). "Automatic calibration of a conceptual rainfall-runoff model using multiple objectives". *J. Hydrol.*, 235(3–4), 276–288.

Melching, C.S. and Yoon, C. G. (1996). "Key sources of uncertainty in QUAL2E model of Passaic River". *J. Water Resour. Plan. Manage.*, 122(2), 105–113.

NRC. (2001). *Assessing the TMDL approach to water quality management*, National Research Council, National Academy Press, Washington, D.C.

NRC. (2004). *Confronting the nation's water problems: The role of research*, National Research Council, The National Academies Press, Washington, D.C.

OMB (Office of Management and Budget) (2003). *Circular A-4*. Washington, DC: OMB.

Oreskes, N. (2003). "The role of quantitative models in science". *In Models in ecosystems science*, Canham, C.D., Cole, J.J. and Lauenroth, W.K., ed., Princeton University Press, NJ, 13–31.

Reichle, R.H., Walker, J.P., Koster, R.D. and Houser, P.R. (2002). "Extended versus ensemble Kalman filtering for land data assimilation". *J. Hydrometeorol.*, 3, 728–740.

Reckhow, K.H. (1994a). "Forum: Importance of scientific uncertainty in decision making". *Environ. Manag.*, 18(2): 161–166.

Reckhow, K.H. (1994b). "Water quality simulation modeling uncertainty analysis for risk assessment and decision making". *Ecol. Modeling.*, 72, 1–20.

Samanta, S., Mackay, D.S., Clayton, M.K., Kruger, E.L. and Ewers, B.E. (2007). "Bayesian analysis for uncertainty estimation of a canopy transpiration model". *Water Resour. Res.*, 43, W04424, doi:10.1029/2006WR005028.

Schnoor, J.L. (1996). *Environmental modeling: Fate and transport of pollutants in water, air and soil*, John Wiley & Sons, NY.

Schulz, K., Beven, K. and Huwe, B. (1999). "Equifinality and the problem of robust calibration in nitrogen budget simulations". *Soil Sci. Soc. Am. J.*, 63,1934–1941.

Shirmohammadi, A., Chaubey, I., Harmel, R.D., Bosch, D.D., Muñoz-Carpena, R., Dharmasri, C., Sexton, A., Arabi, M., Wolfe, M.L., Frankenberger, J., Graff, C. and Shorabi, T.M. (2006). "Uncertainty in TMDL models". *Trans. ASABE*, 49(4): 1033–1049.

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

Stephens, D.B. (2008). "Recent trends in hydrogeology and environmental consulting and perspective on maturing of hydrogeology profession". *J. Hydrol. Engrng.*, 13(1), 20–27.

Thiemann, M., Trosset, M., Gupta, H. and Sorooshian, S. (2001). "Bayesian recursive parameter estimation for hydrologic models". *Water resour. Res.*, 37(10), 2521–2535.

U.S. Environmental Protection Agency. (1997). "Guidelines for preparation of the comprehensive sate water quality assessments", Office of Water, Washington, D.C.

U.S. Environmental Protection Agency. (1999). "Protocols for developing nutrient TMDLs", *Rep.*, EPA-841-B-99-007, Office of Water, Washington, D.C.

U.S. Environmental Protection Agency. (2000). "Nutrient Criteria technical Guidance manual—Lakes and reservoirs", *Rep.*, EPA-822-B-00-001, Office of Water, Washington, D.C.

U.S. Environmental Protection Agency. (2004). "An Examination of EPA risk assessment principles and practices", *Rep.*, EPA-100-B-04-001, Office of the Science Advisor, Washington, D.C.

van der Perk, M. and Bierkens, M.F.P. (1997). "The identifiability of parameters in a water quality model of the Biebrza, Poland". *J. Hydrol.*, 200: 307–322.

Vollenweider, R.A. and Kerekes, J.J. (1980). "Background and summary results of the OECD cooperative program on eutrophication". In *International Symposium on Inland Waters and Lake Restoration*. EPA 440/5-81-010. U.S. Environmental Protection Agency, Washington, DC.

Vrugt, J.A., Gupta, H.V., Bastidas, L.A., Bouten, W. and Sorooshian, S. (2003a). "Effective and efficient algorithm for multiobjective optimization of hydrologic models". *Water Resour. Res.*, 39(8): 1214, doi:10.1029/2002WR001746.

Vrugt, J.A., Gupta, H.V., Bouten, W. and Sorooshian, S. (2003b). "A shuffled complex evolution metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters". *Water Resour. Res.*, 39(8): 1201, doi:10.1029/2002WR001642.

Warwick, J.J. (1997). "Use of first-order uncertainty analysis to optimize successful stream water quality simulation". *J. Am. Water Resour. Assoc.*, 33(6), 1173–1185.

Wu, J., Jones, K.B., Li,H. and Loucks, O.L. (2006). *Scaling and uncertainty analysis in ecology, methods and applications*, Springer, The Netherlands.

Yan, J. and Hann, C.T. (1991). "Multiobjective parameter estimation for hydrologic models- Weighting errors". *Trans. Am. Soc. Agric. Eng.*, 34, 135–141.

Zhang, H.R. and Yu, S.L. (2004). "Applying the first-order error analysis in determining the margin of safety for Total Maximum Daily Load computations". *J. Environ. Engrng.*, 130(6), 664–673.

Zheng, Y. and Keller, A.A. (2007). "Uncertainty assessment in watershed-scale water quality modeling and management: 1. Framework and application of generalized likelihood uncertainty estimation (GLUE) approach". *Water Resour. Res.*, 43, W08407, doi:10.1029/2006WR005345, 1–13.