WATER RESOURCES RESEARCH, VOL. 44, W01420, doi:10.1029/2007WR005862, 2008

Addressing equifinality and uncertainty in eutrophication models

George B. Arhonditsis,[1] Gurbir Perhar,[1] Weitao Zhang,[1] Evangelia Massos,[1] Molu Shi,[1] and Argho Das[1]

Received 2 January 2007; revised 17 August 2007; accepted 19 September 2007; published 17 January 2008.

[1] Large simulation models of eutrophication processes are commonly used to aid scientific understanding and to guide management decisions. Confidence in models for these purposes depends on uncertainty in model equations (structural uncertainty) and on effects of input uncertainties (model parameters, initial conditions, and forcing functions) on model outputs. Our objective herein is to illustrate two strategies, a generalized likelihood uncertainty estimation (GLUE) approach combined with a simple Monte Carlo sampling scheme and a Bayesian methodological framework along with Markov Chain Monte Carlo (MCMC) simulations, for elucidating the propagation of uncertainty in the high-dimensional parameter spaces of mechanistic eutrophication models. We examine the ability of the two approaches to offer insights into the degree of information about model inputs that the data contain, to quantify the correlation structure among parameter estimates, and to obtain predictions along with uncertainty bounds for modeled output variables. Our analysis is based on a four-state-variable (phosphate-detritus-phytoplankton-zooplankton) model and the mesotrophic Lake Washington (Washington State, United States) as a case study. Scientific knowledge, expert judgment, and observational data were used to formulate prior probability distributions and characterize the uncertainty pertaining to 14 model parameters. Despite the conceptual differences for addressing model equifinality, that is, wide ranges of parameter values subject to complex multivariate relationships that result in plausible observed behaviors and produce equivalently accurate predictions, we found that the two strategies provided fairly consistent estimates of the posterior parameter correlation structure and output uncertainty. Nonetheless, our analysis also shows that MCMC can more efficiently quantify the joint probability distribution of model parameters and make inference about this distribution. The latter finding can be explained by the basic idea underlying the MCMC methodology, that is, the configuration of a Markov process whose stationary distribution approximates the joint posterior distribution of all the stochastic model nodes; as a result, Monte Carlo samples are not drawn from the prior parameter space, and problems of wide or highly correlated prior distributions can be overcome. Finally, our study stresses the lack of perfect simulators of natural system dynamics and introduces two statistical formulations that can explicitly account for the discrepancy between mathematical models and environmental systems.

Citation: Arhonditsis, G. B., G. Perhar, W. Zhang, E. Massos, M. Shi, and A. Das (2008), Addressing equifinality and uncertainty in eutrophication models, *Water Resour. Res.*, *44*, W01420, doi:10.1029/2007WR005862.

> If our science is to be meaningful, we should aim to communicate the limitations of the predictions we make in ways that are useful to the wider community. This, in itself, cannot be divorced from the wider socio-political context. . .
>
> *Pappenberger and Beven* [2006, p. 6]

1. Introduction

[2] Mechanistic models are an attractive tool that can be particularly useful for assisting water quality management. An appealing feature for their extensive use is their role as "information integrators" [*Spear*, 1997], that is, their ability to synthesize among different types of information that reflect our existing knowledge/best understanding of the ecosystem functioning. Their main foundation consists of causal mechanisms, complex interrelationships, direct and indirect paths in ecological structures that are mathematically depicted in the form of nonlinear differential equations. Then, any scientific knowledge, expert judgment, and experimental data can be used to assign realistic ranges and determine the relative likelihood of the different values of ecologically meaningful parameters (e.g., chemical processes, biological rates, and partition coefficients). Most importantly, the model endpoints (state variables) usually coincide with routinely monitored environmental variables, which in turn are considered reliable surrogates of the physics, chemistry and biology of the aquatic ecosystem under

[1]Department of Physical and Environmental Sciences, University of Toronto, Toronto, Ontario, Canada.

0043-1397/08/2007WR005862$09.00

investigation. The latter attribute also provides the theoretical underpinning for what is called ''model calibration''; the procedure by which the modelers attempt to minimize the discrepancy between model outputs and observed data by adjusting model parameters [*Jorgensen and Bendoricchio*, 2001]. An implicit assumption for this practice is that if a mathematical model fits well the observed data, then it can be considered an accurate representation of the natural system and can be effectively used for projecting future responses under alternative management schemes [*Arhonditsis and Brett*, 2005a].

[3] Although the premise for ''predictive capability founded on mechanistic understanding'' is tenable, it quickly became clear that there is high degree of model structure and input uncertainty (parameters, initial conditions, forcing functions) resulting in considerable controversy regarding their usefulness as management tools [*Reckhow*, 1994]. This uncertainty is not surprising because all models are simplistic representations of the aquatic systems and even the most well studied ecological processes can be mathematically described by a variety of relationships that entail different assumptions and complexity levels, for example, Monod and Variable-Internal Stores (VIS) models for modeling the phytoplankton uptake of nutrients from the water column and their conversion into biomass [*Grover*, 1991]. Water quality data are also scarce or highly variable, so individual equations that are approximately correct in ideal (controlled laboratory) conditions may not collectively yield an accurate picture of ecosystem behavior. Moreover, the ''chimera'' of a reductionistic description of natural system dynamics accentuates the disparity between what ideally we want to learn and what can realistically be observed, and thus it is often impossible to impose quantitative (or even qualitative) constraints of what should be considered ''behavioral'' simulation [*Beck*, 1987]. By acknowledging the uncertainty (error) underlying both model structures and data, we explicitly recognize that the search for a single set of parameter values (global optimum) that reproduces the real world patterns is not a reasonable expectation [*Reichert and Omlin*, 1997]. Rather, the only defensible strategy is the assessment of the likelihood of different input factors (model structures/parameter sets) being acceptable simulators of the natural system, the so-called ''model equifinality'' [*Beven*, 1993].

[4] Uncertainty analysis of mechanistic models has received considerable attention in the aquatic ecosystem research and there have been several attempts to rigorously address issues pertaining to structural and parametric error [*Spear and Hornberger*, 1980; *Dilks et al.*, 1992; *Omlin and Reichert*, 1999; *Brun et al.*, 2001; *Reichert et al.*, 2002]. Nonetheless, a recent meta-analysis showed that the large majority of the aquatic mechanistic biogeochemical models published over the last decade did not properly assess prediction error; aquatic mechanistic modelers are still reluctant to embrace uncertainty analysis techniques and assess the reliability of the critical planning information generated by the models [*Arhonditsis and Brett*, 2004]. Thorough quantification of model sensitivity to parameters, forcing functions and state variable submodels was only reported in 27.5% of the studies, while 45.1% of the published models did not report any results of uncertainty/sensitivity analysis. Ironically, the identifiability problem in the context of eutrophication modeling and management was firstly discussed nearly three decades ago, when the *Hornberger and Spear* [1981] study advocated the use of regionalized estimation techniques that provide parameter distributions instead of single values (point estimates) as an antidote to the lack of comprehensive data sets. Despite the compelling arguments for considering uncertainty analysis as an integral part of the modeling endeavor, a recent paper by *Pappenberger and Beven* [2006] identified several reasons why the modeling community is still oblivious of its importance and also argued that none of these points is sound. Furthermore, the same study discussed the issues involved in developing a ''Code of Practice'' that aims to offer guidance about how we should conduct uncertainty analysis in any modeling exercise, and the sixth issue of the proposed code, that is, choice of uncertainty estimation methodology, provides the impetus of our study.

[5] The main objective of this study is to illustrate two strategies, a generalized likelihood uncertainty estimation (GLUE) approach combined with a simple Monte Carlo sampling scheme and a Bayesian methodological framework along with Markov Chain Monte Carlo (MCMC) simulations, for assessing the propagation of uncertainty in the high-dimensional parameter spaces of mechanistic eutrophication models. We examine the ability of the two strategies to offer insights into the degree of information the data contain about model inputs, quantify the dependence structure among parameter estimates, and obtain predictions along with uncertainty bounds for modeled output variables. Our goal is to illuminate technical aspects of the two uncertainty analysis approaches that can be particularly useful for eutrophication management. Our illustration is based on a four-state-variable (phosphate-detritus-phytoplankton-zooplankton) model that provides a realistic platform for testing the ability of the competing techniques to explore multidimensional parameter spaces while conforming to the principle of parsimony.

2. Methods

2.1. Model Description

[6] The effectiveness of the two uncertainty analysis approaches was examined using as a case study Lake Washington; the second largest natural lake in Washington State and one of the best documented cases of successful restoration by sewage diversion [*Edmondson*, 1994]. Lake Washington is a mesotrophic system with limnological processes strongly dominated by a recurrent diatom bloom, which occurs during March and April with epilimnetic chlorophyll concentration peaks on average at 10 μg/L, which is 3.2 times higher than the summer concentrations when the system is phosphorus limited [*Arhonditsis et al.*, 2003]. The data set used for model calibration was based on a recent (1994–2003), sampling program carried out by King County/Metro (available at http://dnr.metrokc.gov/wlr/waterres/lakes/LakeWashington.htm). Detailed description of the sampling network, the analytical methods used along with the data analysis is provided elsewhere [*Arhonditsis et al.*, 2003; *Arhonditsis and Brett*, 2005b]. The basic conceptual design of our model builds upon the results of a recent modeling study [*Arhonditsis and Brett*, 2005a, 2005b], and considers the basic ecological processes underlying plank-

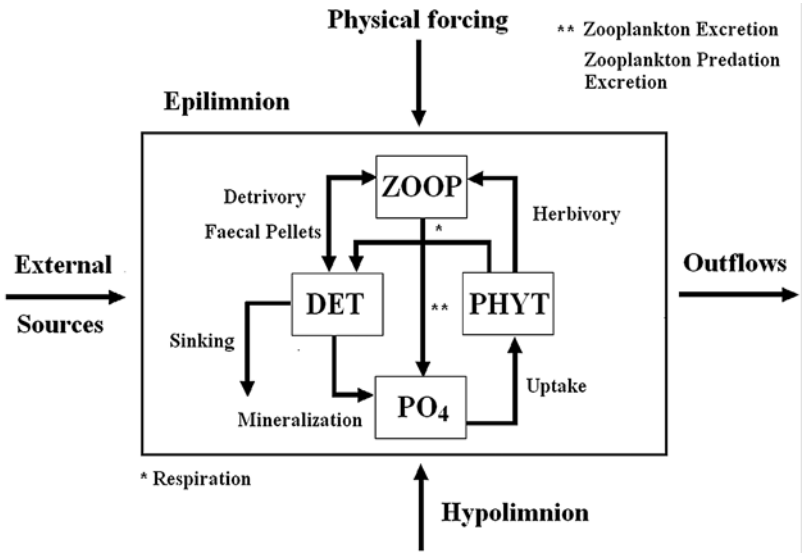

Figure 1. The phosphate (PO_4)–detritus (DET)–phytoplankton (PHYT)–zooplankton (ZOOP) model used for reproducing the Lake Washington dynamics. Arrows indicate flows of matter through the system. System equations and parameter definitions are provided in Appendix A and Table 1.

ton dynamics in the Lake Washington epilimnion. We developed a zero-dimensional (single compartment) model that considers the interplay between the four state variables: phosphate, phytoplankton, zooplankton, and detritus (Figure 1). The mathematical description of the eutrophication model and the definition of the model parameters can be found in Appendix A and Table 1, respectively. The simulation model was solved numerically using the fourth-order Runge-Kutta method with a time step of 1 d.

2.1.1. Phytoplankton

[7] The equation for phytoplankton biomass considers phytoplankton production and losses due to basal metabolism, settling and herbivorous zooplankton grazing. The combined effects of the seasonal cycle of light and temperature (i.e., average physical conditions over the study period) on phytoplankton are described by a trigonometric function $\sigma_{(t)}$. Phytoplankton sinks out of the epilimnion at a constant rate, while phosphorus limitation on phytoplankton growth follows the Michaelis-Menten kinetics. The basal metabolism includes all internal processes that decrease algal biomass (respiration, excretion) as well as natural mortality.

2.1.2. Zooplankton

[8] Zooplankton grazing and losses due to natural mortality/consumption by higher predators are the main two terms in the zooplankton biomass equation, while the specific parameterization used mainly represents *Daphnia* dynamics, that is, the dominant member of the Lake Washington

Table 1. Parameter Definitions of the Eutrophication Model[a]

Parameter	Symbol	Unit
Maximum phytoplankton (PHYT) growth rate	a*	d^{-1}
Higher predation on zooplankton (ZOOP)	d*	d^{-1}
Half-saturation constant for predation	*pred**	mg C m^{-3}
Half-saturation constant for PO_4 uptake	e*	mg P m^{-3}
Cross-thermocline exchange rate	k*	d^{-1}
Phytoplankton respiration rate	r*	d^{-1}
Phytoplankton sinking loss rate	s*	d^{-1}
Phosphorus to carbon ratio for phytoplankton	P/C_{phyto}	0.015 mg P (mg C)$^{-1}$
Phosphorus to carbon ratio for zooplankton	P/C_{zoop}	0.029 mg P (mg C)$^{-1}$
Shape parameter for the trigonometric functions $\sigma_{(t)}$ and $\sigma_{(tz)}$	ε	0.9
Zooplankton growth efficiency	a*	
Zooplankton excretion fraction	β*	
Regeneration of zooplankton predation excretion	γ*	
Maximum zooplankton grazing rate	λ*	d^{-1}
Zooplankton grazing half-saturation coefficient	μ*	mg P m^{-3}
Relative zooplankton preference for detritus compared to phytoplankton	ω	1
Detritus (DET) remineralization rate	φ*	d^{-1}
Detritus sinking rate	ψ*	d^{-1}

[a]The asterisks indicate parameters used during the uncertainty analysis of the model.

zooplankton community [*Arhonditsis and Brett*, 2005a]. Zooplankton has two alternative food sources (phytoplankton and detritus) of equal palatability. A fraction of zooplankton grazing is assimilated and fuels growth, while both herbivory and detrivory were formulated using the Holling Type III function. A sigmoid closure term was selected to represent a "switchable" type of predator behavior controlled by a prey threshold concentration [see *Edwards and Yool*, 2000]. The effects of temperature on zooplankton metabolic activities were modeled by a trigonometric function similar to the one used for phytoplankton. The observed lagged *Daphnia* response (≈30 d) during the spring bloom was represented by a phase shift of −0.5 radians.

2.1.3. Phosphate

[9] The phosphate equation considers the phytoplankton uptake, the proportion of the zooplankton excretion and mortality/predation that is returned back to the system as dissolved phosphorus. Epilimnetic phosphate levels are also fuelled by the bacteria-mediated mineralization of detritus, and are subject to seasonally varying diffusive mixing with the hypolimnion.

2.1.4. Detritus

[10] Detritus sinks out of the epilimnion at a constant rate and is transformed to phosphate by the seasonally forced mineralization processes. Phytoplankton respiration and a fraction of the zooplankton growth that represents the fecal pellets also contribute to the detritus pool.

2.1.5. Boundary Conditions

[11] The dissolved phosphorus and detritus (mainly representing the inflows of particulate phosphorus) loadings from the watershed, the hypolimnetic phosphate concentrations and the outflows to the Lake Union Ship Canal were described by sinusoidal functions fitted to data from the local streams [*Brett et al.*, 2005], the King County/Metro sampling program and the H. H. Chittenden Locks [*Arhonditsis and Brett*, 2005a].

2.2. Generalized Likelihood Uncertainty Estimation

[12] The generalized likelihood uncertainty estimation is an extension of the binary "acceptance/rejection" system of behavioral/nonbehavioral simulations of the original Regionalized Sensitivity Analysis [*Hornberger and Spear*, 1981]. The GLUE methodology uses likelihood measures to assign different levels of confidence (weighting) to different parameter sets, reflecting their ability to acceptably reproduce "non-error-free" observations from the environmental system [*Beven and Binley*, 1992; *Zak and Beven*, 1999; *Page et al.*, 2004; *Beven*, 2006]. Unlike other uncertainty analysis techniques, the term likelihood has a very broad meaning and is specified as any measure of goodness-of-fit that can be used to compare observed responses with model predictions [*Zak et al.*, 1997]. As a result, a wide variety of likelihood functions can be found in the GLUE literature, for example, likelihood measures based on the sum of squared errors [*Beven and Binley*, 1992; *Sorooshian and Gupta*, 1995; *Freer et al.*, 1997], fuzzy measures [*Franks et al.*, 1998; *Page et al.*, 2004] or even qualitative measures for model evaluation [*Beven*, 2001]. The GLUE procedure requires a large number of Monte Carlo model runs sampled from (usually) uniform distributions across a plausible range of each parameter. The behavioral runs are selected on the basis of a subjectively chosen threshold of the likelihood function and are rescaled so that their cumulative total is 1.0. The weighting assigned to the retained behavioral runs is propagated to the model output and forms a likelihood-weighted cumulative distribution of the predicted variable(s), which are then used for estimating the prediction uncertainty ranges [*Beven and Binley*, 1992].

[13] GLUE also has the ability to update likelihood weights (and thus predictive uncertainty) by successive application of the Bayes' Theorem, as additional data become available. The refinement of the predictive uncertainty can be assessed each time the likelihood function is updated by the use of appropriate quantitative measures (e.g., the probabilistic Shannon Entropy measure and the U-uncertainty); see the relevant discussion by *Beven and Binley* [1992]. However, unless the combination of a "likelihood" function with a threshold criterion corresponds to a well-defined probability distribution that directly connects the data with model input and output parameters, GLUE does not have a clear Bayesian interpretation [*Engeland and Gottschalk*, 2002; *Hong et al.*, 2005]. For example, the 95% uncertainty bounds resulting from GLUE can have different statistical interpretation than the 95% credible intervals [*Lamb et al.*, 1998]. In addition, the GLUE procedure is mainly based on the parameter-set ability to produce behavioral simulations, and therefore it is difficult to extract information regarding the individual parameter effects on the model response. We know, for example, from past studies that the set of modal values of the marginal distributions may not be itself behavioral [*Beven*, 2006]. Even though the use of scattergrams (projections of the sampled high-dimensional response surface onto single parameter dimensions) can give insights, they cannot fully reflect the complexity of the response surface [*Page et al.*, 2004]. Finally, it should also be noted that depending on the likelihood measure and the behavioral/nonbehavioral threshold used, GLUE can be particularly inefficient in sampling acceptable runs, for example, the MAGIC application by *Page et al.* [2004] classified as acceptable 7200 out of eleven million runs (0.066%).

2.3. Monte Carlo Sampling Specification of Likelihood Weights

[14] For the GLUE illustration, the uncertainty associated with the initial conditions was accommodated by using the same initial values (mean January concentrations over the study period) for all the Monte Carlo runs and a 10-a simulation period which was sufficient time to reach an equilibrium state (i.e., reproduce similar annual cycles) or to collapse (zero, negative values or approach infinity). To further retain the dimensions of the space examined to a manageable level, we did not consider the error underlying the input boundary conditions. Furthermore, since 4 parameters out of 18 are already known from empirical data and literature information, the total number of unknown parameters is 14 (Table 1). To maximize the efficiency (acceptance rates) of our Monte Carlo sampling, we implemented a modest pruning of the initial parameter space using 25,000 random sets sampled from 14 uniform distributions. The ranges for this 14-dimensional hypercube were based on the identified minimum and maximum parameter values from the pertinent literature [*Arhonditsis and Brett*, 2005a, and references therein]. This exploratory analysis resulted in a 5–20% reduction of the ranges of the univariate marginal distributions, although behavioral simulations were found

Table 2. Prior and GLUE Estimates of the Mean Values and Standard Deviations of the Eutrophication Model Parameters[a]

	Prior		GLUE	
Parameter	Mean	SD	Mean	SD
a	1.297	0.221	1.071	0.161
d	0.175	0.016	0.174	0.016
pred	57.55	10.06	59.81	10.04
e	15.03	4.500	18.59	3.549
k	0.035	0.007	0.034	0.007
r	0.095	0.025	0.098	0.025
s	0.055	0.020	0.044	0.009
a	0.401	0.090	0.425	0.077
β	0.325	0.078	0.328	0.074
γ	0.325	0.078	0.332	0.080
λ	0.624	0.078	0.650	0.074
μ	7.016	1.350	6.353	1.217
φ	0.225	0.078	0.238	0.074
ψ	0.080	0.032	0.044	0.024

[a]GLUE, generalized likelihood uncertainty estimation.

over the entire allowable regions for some parameters. The latter finding has been repeatedly reported in the modeling literature [e.g., *Spear et al.*, 1994] and underscores the need for caution when modelers attempt to truncate the parameter space. In this study, our decisions were also guided from existing knowledge and past experience regarding the magnitudes of the various ecological processes in a mesotrophic environment and the parameter regions that increase the likelihood to reproduce the observed Lake Washington patterns [*Arhonditsis and Brett*, 2005b].

[15] Using the same rationale, we also opted for informative parameter distributions for the GLUE analysis, and their characterization was based on the assignment of normal probability densities to the adjusted parameter ranges (Table 2). The prior distribution of each model input parameter was determined independently of other parameters for two reasons: (1) there was no consistent literature information regarding correlations among model parameters [*Di Toro and van Straten*, 1979], and (2) the independence assumption results in a multivariate prior distribution that tends to be more spread out than one that attempts to incorporate dependence explicitly. We used the latter feature to provide a type of prior robustness in the analysis that compensated for the compromises made during the truncation of the initial parameter space; that is, instead of a wide hypercube or a narrow hyperellipsoid, we focused on a relatively flat subregion where there was evidence (e.g., scientific knowledge, expert judgment, and observational data) for more realistic representation of the system dynamics. Furthermore, the assumption of a priori independence does not imply independence a posteriori and some information on the parameter covariance in fitting the available observations will be unraveled in the posterior distribution of likelihood measures following the conditioning of the parameter sets within the GLUE procedure (see section 3). After the formation of the density function that describes the joint probabilities of the 14 parameters, we generated a number of 60,000 random sets, which then provided the input for our eutrophication model.

[16] As previously mentioned, the definition of the likelihood measure that assesses the goodness-of-fit of model outputs to the observed data is a critical step in the GLUE framework and the uncertainty predictions can be strongly influenced by that definition [*Ratto et al.*, 2001]. In this study, we employed the following likelihood measure:

$$L_j(\theta^k|Y_j) = \frac{1}{\sigma_j^{k2}}, \tag{1}$$

where

$$\sigma_j^{k2} = \frac{1}{2n}\sum_{i=1}^{n}\left(Y_{obs(ij)} - Y_{pred(ij)}^{k}\right)^2 \quad j \in \{PO4, det, chla, zoop\} \tag{2}$$

is the mean squared difference between predicted and observed values for the j state variable from the kth Monte Carlo simulation, and n is the number of observations (twelve average monthly values). The four L_j values were multiplied to give a combined likelihood measure for each Monte Carlo run. After the selection of the behavioral simulations (see section 3), the likelihood measures were rescaled from 0 to 1 and these rescaled weights were then used to compute the posterior parameter (i.e., mean, variance, and correlation coefficients) and state variable (i.e., weighted mean predicted monthly values, weighted variance, and selected percentiles from the cumulative likelihood distributions) statistics. We also examined the consistency of the results obtained from different likelihood measures (behavioral runs, posterior parameter statistics, predictive uncertainty) by considering two additional measures of fit, that is, the relative error [RE = Σ| Observed values − Predicted values|/ΣObserved values] and the modeling efficiency [MEF = 1 − Σ(Predicted values − Observed values)^2/Σ(Observed values − Observation average)^2] [*Stow et al.*, 2003].

2.4. Bayesian Methodological Framework

[17] Our presentation examines three statistical formulations that aim to combine field observations and simulation model outputs to update the uncertainty of model parameters, determine their correlation structure, and then use the calibrated model to give predictions (along with uncertainty bounds) of the natural system dynamics. The three approaches can be distinguished by the following assumptions: (1) the eutrophication model is a perfect simulator of the environmental system (i.e., the difference between model and lake dynamics was assumed to be caused only by the observation error) (Model 1), (2) the eutrophication model is an imperfect simulator of the environmental system and the model discrepancy is invariant with the input conditions (i.e., the difference between model and lake dynamics was assumed to be constant over the annual cycle for each state variable) (Model 2), and (3) the eutrophication model is an imperfect simulator of the environmental system and the model discrepancy varies with the input conditions (i.e., there is seasonally varying discrepancy between model and lake dynamics for each state variable) (Model 3).

2.4.1. Model 1

[18] The first statistical formulation is based on the assumption that our model perfectly describes the dynamics of the environmental system and the observations y for the four state variables is given by

$$y_i = f(\theta, x_i, y_0) + \varepsilon_i, \; i = 1, 2, 3, \ldots, n, \tag{3}$$

where $f(\theta, x_i, y_0)$ denotes the eutrophication model, x_i is a vector of time-dependent control variables (e.g., boundary conditions and forcing functions) describing the environmental conditions, the vector θ is a time-independent set of the calibration model parameters (i.e., the 14 parameters in Table 1), y_0 corresponds to the concentrations of the four state variables at the initial time point t_0, and ε_i denotes the observation (measurement) error that is usually assumed to be independent and identically distributed following a Gaussian distribution. The observed Lake Washington patterns provide evidence of a multiplicative measurement error [*Arhonditsis et al.*, 2003, Figure 2], and thus we assumed the standard deviation to be proportional to the average monthly values for each state variable [*Van Oijen et al.*, 2005].

[19] On the basis of the previous assumptions, the likelihood function that evaluates how well the simulation model is able to reproduce the observed data y at each value of θ, is given by

$$p(y|f(\theta,x,y_0)) = \prod_{j=1}^{m} (2\pi)^{-n/2}\left|\Sigma_{\varepsilon j}\right|^{-1/2} \cdot \exp\left[-\frac{1}{2}\left[y_j - f_j(\theta,x,y_0)\right]^T \cdot \Sigma_{\varepsilon j}^{-1}\left[y_j - f_j(\theta,x,y_0)\right]\right], \quad (4)$$

where m and n correspond to the number of state variables (m = 4) and the number of observations in time used to calibrate the model (n = 12 average monthly values), respectively; $y_j = [y_{1j}, \ldots, y_{nj}]^T$ and $f_j(\theta, x, y_0) = [f_j(\theta, \boldsymbol{x}_i, y_0), \ldots, f_j(\theta, x_n, y_0)]^T$ correspond to the vectors of the field observations and model predictions for the state variable j; and $\Sigma_{\varepsilon j} = I_n \cdot (0.15)^2 \cdot y_j^T \cdot y_j$ where the factor 0.15 reflects our assumption that the monthly standard deviations are 15% of the mean monthly values; a fraction that comprises both analytical error and interannual variability at the deeper (middle) sections of the lake. In the context of the Bayesian statistical inference, the posterior density of the parameters θ and the initial conditions of the four state variables y_0 given the observed data y is defined as

$$p(\theta,y_0|y) = \frac{p(y|f(\theta,x,y_0))p(\theta)p(y_0)}{\iint p(y|f(\theta,x,y_0))p(\theta)p(y_0)d\theta dy_0} \propto p(y|f(\theta,x,y_0))p(\theta)p(y_0), \quad (5)$$

where $p(\theta)$ is the prior density of the model parameters θ and $p(y_0)$ is the prior density of the initial conditions of the four state variables y_0. In a similar way to the measurement errors, the characterization of the prior density $p(y_0)$ was based on the assumption of a Gaussian distribution with a mean value derived from the January monthly averages during the study period and standard deviation that was 15% of the mean value for each state variable j; a fraction that comprises both analytical error, interannual variability at the deeper (middle) sections of the lake, and difference between the starting date of the simulation period (1 January) and the usual sampling dates of the King County/Metro sampling program. Thus the resulting posterior distribution for θ and y_0 is given by

$$p(\theta,y_0|y) \propto \prod_{j=1}^{m} (2\pi)^{-n/2}\left|\Sigma_{\varepsilon j}\right|^{-1/2} \cdot \exp\left[-\frac{1}{2}\left[y_j - f_j(\theta,x,y_0)\right]^T \Sigma_{\varepsilon j}^{-1}\left[y_j - f_j(\theta,x,y_0)\right]\right] \times (2\pi)^{-l/2}|\Sigma_\theta|^{-1/2}\prod_{k=1}^{l}\frac{1}{\theta_k} \cdot \exp\left[-\frac{1}{2}[\log\theta - \theta_0]^T\Sigma_\theta^{-1}[\log\theta - \theta_0]\right] \times (2\pi)^{-m/2} \cdot \left|\Sigma_{y0}\right|^{-1/2}\exp\left[-\frac{1}{2}[y_0 - y_{0m}]^T\Sigma_{y0}^{-1}[y_0 - y_{0m}]\right], \quad (6)$$

where l is the number of the model parameters θ used for the model calibration (l = 14); θ_0 denotes the vector of the mean values of θ (logarithmic scale); $\Sigma_\theta = I_l \cdot \sigma_\theta^T \cdot \sigma_\theta$ and $\sigma_\theta = [\sigma_{\theta 1}, \ldots, \sigma_{\theta l}]^T$ corresponds to the vector of the shape parameters of the l lognormal distributions (standard deviation of log θ); the vector $y_{0m} = [y_{11}, \ldots, y_{14}]^T$ corresponds to the average values of the four state variables observed in January during the study period (1994–2003); and $\Sigma_{y0} = I_m \cdot (0.15)^2 \cdot y_{0m}^T \cdot y_{0m}$.

2.4.2. Model 2

[20] An augmentation of the previous statistical formulation explicitly recognizes that the model imperfectly represents the dynamics of the environmental system. In this case, an observation i for the state variables j, y_{ij}, can be described as

$$y_{ij} = f(\theta,x_i,y_0) + \delta_j + \varepsilon_{ij}, \quad i = 1,2,3,\ldots..n \text{ and } j = 1,\ldots,m, \quad (7)$$

where the stochastic term δ_j accounts for the discrepancy between the model $f(\theta, x, y_0)$ and the natural system, which is assumed to be invariant with the input conditions x (i.e., the difference between model and lake dynamics was assumed to be constant over the annual cycle for each state variable). With this assumption, the likelihood function will be

$$p(y|f(\theta,x,y_0)) = \prod_{j=1}^{m} (2\pi)^{-n/2}\left|\Sigma_{Tj}\right|^{-1/2} \cdot \exp\left[-\frac{1}{2}\left[y_j - f_j(\theta,x,y_0)\right]^T\Sigma_{Tj}^{-1} \cdot \left[y_j - f_j(\theta,x,y_0)\right]\right], \quad (8)$$

$$\Sigma_{Tj} = \Sigma_{\delta j} + \Sigma_{\varepsilon j}, \quad (9)$$

where $\Sigma_{\delta j} = I_n \cdot \boldsymbol{\sigma_j^2}$ corresponds to the additional stochastic term of Model 2; and the prior densities $p(\boldsymbol{\sigma_j^2})$ were based on uniform distributions to overcome some problems caused from

the conjugate inverse-gamma distribution [*Gelman*, 2005]. Thus the resulting posterior distribution for θ, y_0, and σ^2 is

$$\begin{aligned} p(\theta, y_0, \sigma^2|y) \propto & \prod_{j=1}^{m} (2\pi)^{-n/2} \left|\Sigma_{Tj}\right|^{-1/2} \\ & \cdot \exp\left[-\frac{1}{2}\left[y_j - f_j(\theta, x, y_0)\right]^T \Sigma_{Tj}^{-1} \left[y_j - f_j(\theta, x, y_0)\right]\right] \\ & \times (2\pi)^{-l/2} |\Sigma_\theta|^{-1/2} \prod_{k=1}^{l} \frac{1}{\theta_k} \\ & \cdot \exp\left[-\frac{1}{2}\left[\log\theta - \theta_0\right]^T \Sigma_\theta^{-1} \left[\log\theta - \theta_0\right]\right] \\ & \times (2\pi)^{-m/2} \left|\Sigma_{y0}\right|^{-1/2} \\ & \cdot \exp\left[-\frac{1}{2}\left[y_0 - y_{0m}\right]^T \Sigma_{y0}^{-1} \left[y_0 - y_{0m}\right]\right] \\ & \times \prod_{j=1}^{m} \frac{1}{up_j - lo_j}, \end{aligned} \tag{10}$$

where the location parameters *lo* and *up* correspond to the lower and upper limit of the range of the m uniform distributions.

2.4.3. Model 3

[21] The third statistical formulation also explicitly recognizes that the model imperfectly represents the dynamics of the environmental system but now the corresponding stochastic term varies with the input conditions x. In this case, an observation i for the state variables j, y_{ij}, can be described as

$$y_{ij} = f(\theta, x_i, y_0) + \delta_{ij} + \varepsilon_{ij}, \quad i = 1, 2, 3, \ldots . n \text{ and } j = 1, \ldots, m. \tag{11}$$

The modeling for all the previous terms remains unchanged. We also specify a Gaussian first-order random walk model for the discrepancy term δ_{ij} to reflect that these error terms are correlated [*Shaddick and Wakefield*, 2002]. Specifically, the vector $\delta_j = \{\delta_{1j}, \ldots, \delta_{12j}\}$, $j = PO_4$, *det*, *chla*, *zoop*, can be expressed as

$$p\left(\delta_{ij}|\delta_{-ij}, \omega_j^2\right) \sim \begin{cases} N\left(\delta_{i+1j}, \omega_j^2\right) \text{ for } i = 1, \\ N\left(\dfrac{\delta_{i-1j} + \delta_{i+1j}}{2}, \dfrac{\omega_j^2}{2}\right) \text{ for } i = 2, \ldots, 11, \\ N\left(\delta_{i-1j}, \omega_j^2\right) \text{ for } i = 12, \end{cases} \tag{12}$$

where δ_{-ij} denotes all elements of δ_j except the δ_{ij}, ω_j^2 is the conditional variance and the prior densities $p(\omega_j^2)$ were based on conjugate inverse-gamma (0.01, 0.01) distributions [*Gelman et al.*, 1995]. Finally, it is interesting to note that the third statistical formulation has conceptual similarities with the hierarchical dynamic linear modeling (DLM) [*Pole et al.*, 1994] and the Kalman filter [*Meinhold and Singpurwalla*, 1983].

[22] Assessment of the goodness-of-fit between the model predictions and the observed data was based on the posterior predictive p value, that is, the Bayesian counterpart of the classical p value. In brief, the posterior predictive p value is defined as the probability that the replicated data (the posterior predictive distribution) could be more extreme than the observed data. The null hypothesis H_0 (i.e., there are no systematic differences between the simulations and the data) is rejected if the tail-area probability is close to 0.0 or 1.0; whilst the model can be regarded as plausible if the p value is near to 0.5. The "discrepancy variable" chosen for carrying out the posterior predictive model checks was the x^2 test (see also *Gelman et al.* [1996] for a detailed description of the posterior predictive p value). The comparison between the two alternative models was based on the use of the Bayes factor, that is, the posterior odds of one model over the other (assuming the prior probability on either model is 0.5). If M_1 and M_2 denote the two alternative models, the Bayes factor is

$$B_{12} = \frac{pr(y|M_1)}{pr(y|M_2)}. \tag{13}$$

For model comparison purposes, the model likelihood ($pr(y|M_w)$; w = 1, 2) is obtained by integrating over the unknown element (initial conditions, model parameters, error terms) space:

$$pr(y|M_w) = \int pr(y|M_w, \Theta_w)\pi(\Theta_w|M_w)d\Theta_w, \tag{14}$$

where Θ_w is the unknown element vector under model M_w and $\pi(\Theta_w|M_w)$ is the prior density of Θ_w. Using the MCMC method, we can estimate $pr(y|M_w)$ from posterior samples of Θ_w. Letting $\Theta_w^{(i)}$ be samples from the posterior density $pr(\Theta_w|M_w)$, the estimated $pr(y|M_w)$ is

$$\overline{pr(y|M_w)} = \left\{\frac{1}{m}\sum_{i=1}^{m} pr\left(y|M_w, \Theta_k^{(i)}\right)^{-1}\right\}^{-1}, \tag{15}$$

the harmonic mean of the likelihood values [*Kass and Raftery*, 1995].

2.4.4. Markov Chain Monte Carlo

[23] Markov Chain Monte Carlo (MCMC) is a general methodology that provides a convenient means for sampling multidimensional distributions for the purpose of numerical integration. The idea underlying the MCMC implementation in Bayesian inference is to construct a random walk or Markov process whose stationary distribution is $p(f(\theta, x)|y)$ (i.e., the combination of the prior information with the model likelihood under the observed data) and then run the process long enough so that we produce an accurate approximation of the posterior model distribution [*Gilks et al.*, 1998]. There are many methods (e.g., Gibbs sampler) for obtaining sequence of realizations from the posterior model distributions [*Gelfand and Smith*, 1990; *Casella and George*, 1992], but all of them are special cases of the general Metropolis-Hastings algorithm [*Metropolis*, 1953; *Hastings*, 1970].

[24] To generate samples from a target posterior distribution $p(f(\theta, x)|y)$, we need to know that distribution up to a proportional constant and the simplest form of the Metropolis-Hastings algorithm is summarized as follows: (1) select an initial value $\theta_{t=0}$, for which $p(f(\theta_{t=0}, x)|y) > 0$, from a prior distribution $p(\theta)$, (2) generate θ^* from a symmetric distribution (i.e., the chance of generating θ^* given θ^t is the same as generating θ^t given θ^*) and u from

the uniform distribution U(0,1), (3) compute the ratio of the densities,

$$r = \frac{p(\theta^*|y)}{p(\theta^t|y)}, \quad (16)$$

(4) compute the probability of move α as

$$a(\theta_t, \theta^*) = \min[r, 1], \quad (17)$$

(5) if u $\leq \alpha$, set $\theta_{t+1} = \theta^*$; else set $\theta_{t+1} = \theta_t$, and (6) repeat steps 2–5, starting at θ_{t+1}.

[25] This simple but general procedure has several advantages that are particularly useful for environmental modeling; that is, it can efficiently sample multidimensional parameter spaces and can easily handle multivariate outputs as well as large numbers of nuisance parameters [*Gelman et al.*, 1995].

[26] In this study, we used the general normal-proposal Metropolis algorithm as it is implemented in the WinBUGS software (WinBUGS User Manual, version 1.4, D. Spiegelhalter et al., 2003, available at http://www.mrc-bsu.cam.ac.uk/bugs); this algorithm is based on a symmetric normal proposal distribution, whose standard deviation is adjusted over the first 4000 iterations such as the acceptance rate ranges between 20% and 40%. We also used an ordered over-relaxation, which generates multiple samples per iteration and then selects one that is negatively correlated with the current value of each stochastic node [*Neal*, 1998]. The latter option resulted in an increased time per iteration but reduced within-chain correlations. The posterior simulations were based on multiple chains from starting points dispersed around the parameter space [*Steinberg et al.*, 1996]. We found that some of the initial parameter vectors resulted in unstable solutions that tended to infinity, while the presented results are based on two parallel chains with starting points: (1) a vector that consists of the mean values of the prior parameter distributions and (2) a vector that resulted from the optimization of the model with the Fletcher-Reeves conjugate-gradient method [*Chapra and Canale*, 1998]. We used 30,000 iterations and convergence was assessed with the modified Gelman–Rubin convergence statistic [*Brooks and Gelman*, 1998]. The accuracy of the posterior estimates was inspected by assuring that the Monte Carlo error (an estimate of the difference between the mean of the sampled values and the true posterior mean; see the WinBUGS User Manual, version 1.4, D. Spiegelhalter et al., 2003, available at http://www.mrc-bsu.cam.ac.uk/bugs) for all the parameters was less than 5% of the sample standard deviation. Given the advantages of the MCMC methodology, the characterization of the prior density functions were somewhat more realistic (less constrained) relative to those used for the GLUE applications; that is, we assigned lognormal distributions that 95% of their values were lying within the identified ranges for each parameter [*Steinberg et al.*, 1997].

3. Results

3.1. Generalized Likelihood Uncertainty Estimation

[27] The selection of the behavioral runs for the GLUE analysis was based on the overall model likelihood (equation (1)), using as a threshold level the lowest L value that resulted in less than 20% (10 out of 48) and no violations of the bands defined by the monthly averages ±1 and 2 standard deviations (0.15 × monthly average), respectively. We found that 832 runs met this criterion and the corresponding likelihood measures were then rescaled such that the sum of all the likelihood values was equal to 1. The distribution of the rescaled likelihood measures for six of the model parameters (a, e, s, α, μ, and ψ) is shown in Figure 2. These scatterplots represent the projection of the 14-dimensional parameter response surface onto single parameter axes, where it can be seen that good simulations were produced throughout the chosen parameter ranges. Although we were able to locate some parameter subregions with more frequent occurrences of high model performance ($L > 0.004$), the general patterns suggest that the likelihood surface is very complex with many smaller peaks, so that no single global optimum parameter set could be identified (see also section 4). Consequently, individual parameter values are not particularly important in the prediction of lake dynamics if taken outside the context of the values of the other parameters [*Schulz et al.*, 1999]. In this regard, one of the advantages of the GLUE methodology is that the parameters are treated as sets, and thus the effect of such interactions are implicitly reflected in the likelihood value associated with each set [*Zak and Beven*, 1999].

[28] The likelihood-weighted mean values indicate that the majority of the posterior parameter distributions were characterized by minor shifts of their central tendency relative to the prior assigned distributions (Table 2). The only exception was the detritus sinking rate (ψ) with a 45% decrease of the posterior mean (0.044 from 0.08), which was also accompanied by a 24% relative decrease of the respective standard deviation. On the other hand, the maximum phytoplankton growth rate (a), the half-saturation constant for PO4 uptake (e), the phytoplankton sinking rate (s), the zooplankton assimilation efficiency (a), and the detritus sinking rate (ψ) showed the most significant relative decrease (14 to 55%) of their standard deviations, while the knowledge gained for the rest parameters was fairly minimal ($\leq$10% relative reduction). The behavioral subset can also be used to determine the correlation structure among the model parameters (Table 3), although it must be noted that the global correlation coefficients (derived from all the behavioral samples) can be misleading because the local interaction between parameters are more important; for example, there might be cases in which positive interactions in one part of the space can produce a behavioral model, while elsewhere there might be a negative interaction. Despite the aforementioned caution regarding the global correlation coefficient values, some of the manifested relationships have plausible physical explanation. For example, a high maximum phytoplankton growth rate can be balanced by a high half-saturation constant for PO_4 uptake (e) or phytoplankton sinking loss rate (s) to accurately represent the observed epilimnetic phytoplankton dynamics. There were also some relationships that seem counter-intuitive, such as the positive correlation between the detritus mineralization (φ) and sinking rates (ψ). Given that both parameters reflect detritus losses from the system, a negative relationship where the two terms cancel each other out would have seemed more plausible. The positive correlation probably reflects the predominance of other

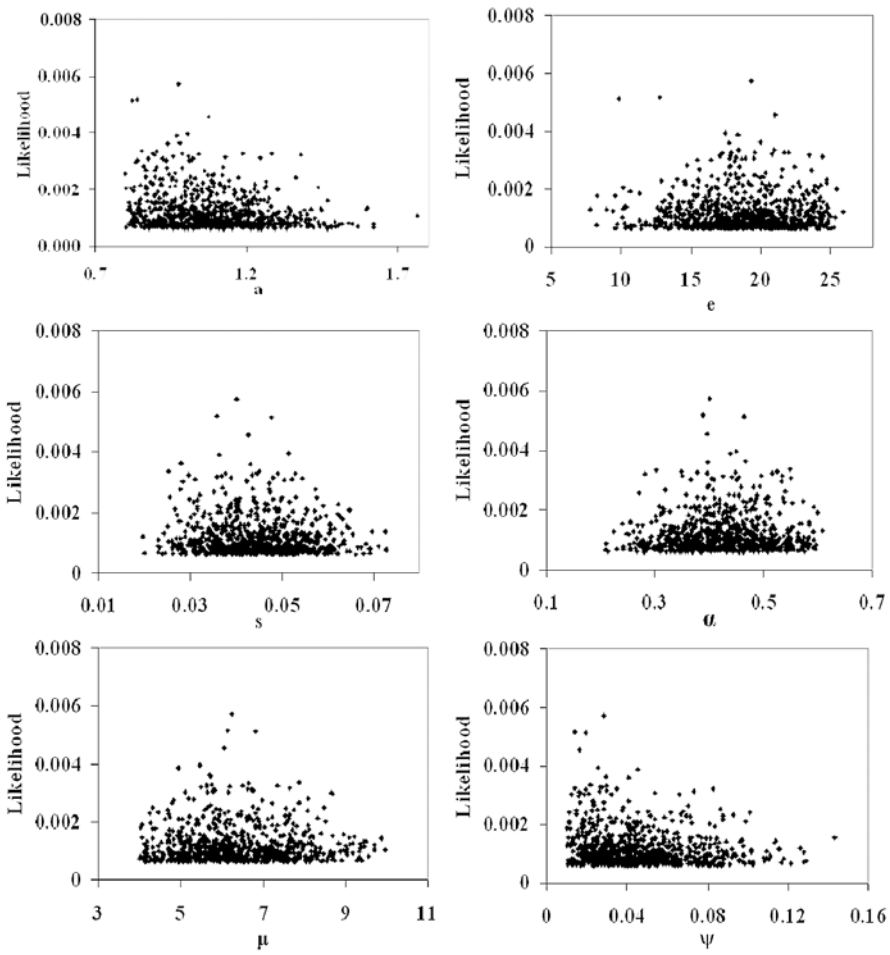

Figure 2. Scatterplots of the likelihood measure $1/\sigma^2$ versus the eutrophication model parameters a, e, s, α, μ, and ψ.

ecological paths considered in our eutrophication model; for example, higher detritus sinking rates require higher detritus mineralization rates to fuel phytoplankton growth and compensate for the zooplankton food deficit, while the detritus pool is fuelled by both the increased phytoplankton respiration and zooplankton grazing and thus the model can still provide a reasonable fit to the observed data.

[29] The comparison between the observed data and the model outputs shows that the majority of the phosphate, chlorophyll a and zooplankton biomass monthly values were included within the 2.5 and 97.5% uncertainty bounds (Figure 3). Nonetheless, the likelihood-weighted mean predicted phytoplankton and zooplankton values underrepresented the observed spring plankton dynamics and approximately half of the modeled total phosphorus concentrations were not bracketed by the respective confidence

Table 3. Correlation Matrix of the Eutrophication Model Parameters Based on GLUE Analysis[a]

	a	d	Pred	e	k	r	s	α	β	γ	λ	μ	φ	Ψ
		−0.062	0.002	**0.471**	−0.008	0.072	**0.362**	−0.061	0.046	0.088	0.013	−0.131	−0.020	−0.168
a			0.084	0.005	−0.018	0.042	−0.066	0.112	0.011	0.062	0.053	−0.053	0.023	−0.051
d				0.025	0.018	−0.081	0.035	−0.085	0.011	0.000	−0.080	0.099	0.155	0.203
Pred					−0.040	−0.147	−0.242	0.027	0.001	−0.013	−0.077	0.038	0.017	0.140
e						0.061	0.106	−0.040	0.043	−0.007	0.041	0.052	0.022	−0.045
k							−0.198	−0.055	0.051	−0.022	−0.042	0.091	−0.030	**−0.266**
s								−0.046	0.074	0.127	0.056	−0.177	0.103	**−0.401**
α									0.023	0.086	−0.094	0.150	0.205	0.157
β										−0.044	0.029	0.002	−0.054	−0.046
γ											0.016	−0.029	0.021	0.080
λ												0.144	0.036	0.101
μ													−0.107	**−0.260**
φ														**0.420**
ψ														

[a]Bold numbers correspond to correlation coefficients with absolute value greater than 0.250.

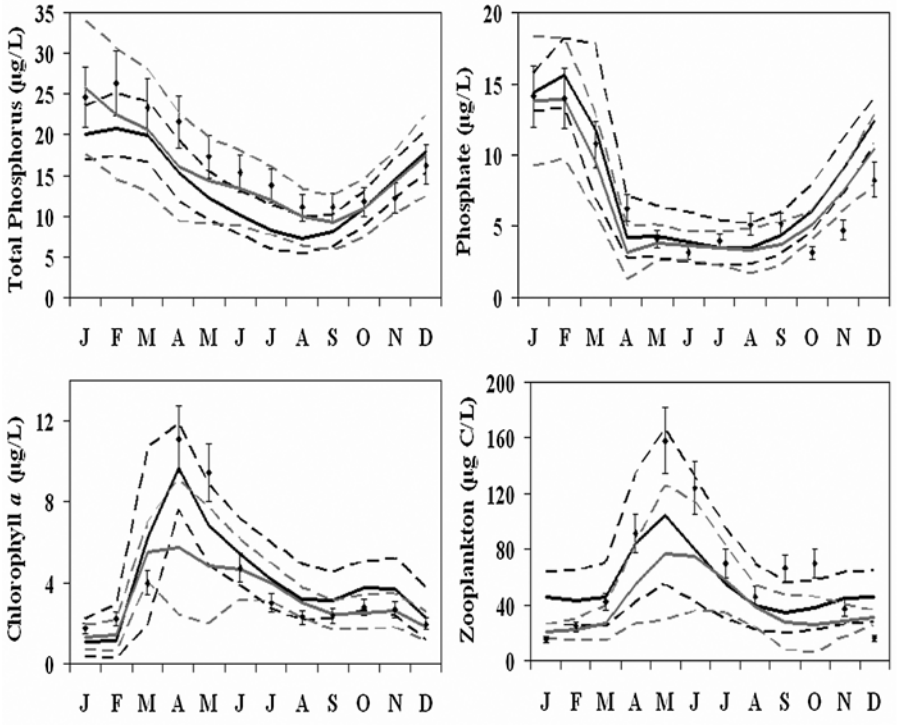

Figure 3. Observed monthly values versus the likelihood-weighted mean predicted values for total phosphorus, phosphate, chlorophyll a (1 g carbon = 20 mg chlorophyll), and zooplankton biomass based on the generalized likelihood uncertainty estimation (GLUE) analysis (black lines). Grey lines correspond to the posterior predictive monthly distributions from Model 1. Dashed lines correspond to the 2.5 and 97.5% uncertainty bounds of the two models. Single dots and the respective black lines correspond to the monthly averages and standard deviations, reflecting the analytical error and the interannual variability in the lake over the 10-a period, 1994–2003. (The modeled total phosphorus concentrations comprise three phosphorus pools, i.e., phosphate, detritus, and phosphorus sequestered in phytoplankton cells.)

limits. Interestingly, the optimization of the model (not presented here) using Powell's direct pattern search and Fletcher-Reeves conjugate gradient method along with a cost function that equally weights the four state variables gave almost the same results and only slightly improved the representations of the mid/late spring plankton dynamics. Thus we hypothesized that the former result is probably associated with the model error structure, while the latter stems from the inaccuracy of the proxy used for the representation of the particulate phosphorus exogenous loadings. Whether the model lack of fit along with the minor shifts of the updated parameter distributions reflect inefficient sampling from misformulated prior parameter distributions or model misspecification (e.g., inadequate model structure and/or inaccurate boundary conditions) was further examined with the MCMC sampling scheme. Finally, the use of alternative likelihood measures, that is, the relative error (threshold value $\leq$ 40%) and the modeling efficiency ($\geq$0.4), provided relatively similar results regarding the posterior parameter statistics and the predictive uncertainty bounds, although we found a 60–80% overlap among the three subsets of behavioral runs derived from the different measures of fit.

3.2. Bayesian Formulations

[30] The two MCMC sequences of the three models converged rapidly (≈5000 iterations) and the statistics reported in this study were based on the last 25,000 draws by keeping every 10th iteration (thin = 10). The uncertainty underlying the values of the 14 model parameters after the MCMC sampling is depicted on the respective marginal posterior distributions (Table 4 and Figure 4). Generally, the standard deviation of the posterior parameter distributions were significantly reduced with the first statistical formulation (Model 1); characteristic examples were the half-saturation constant for predation (*pred*), the cross-thermocline exchange rate (k), the phytoplankton sinking loss rate (s), the detritus mineralization (φ) and sinking rates (ψ) with a relative decrease higher than 80%. On the other hand, the inclusion of the seasonally invariant stochastic term that accounts for the discrepancy between the model and the real system resulted in higher standard deviations for most of the parameters, and, in some cases, our knowledge did not improve relative to what we knew prior to the calibration (e.g., higher predation on zooplankton (d), half-saturation constant for predation (*pred*), and cross-thermocline exchange rate (k)). The latter finding indicates that the discrepancy term in Model 2 mainly improves our knowledge on the natural system dynamics (see the following results) but gives little information regarding the values of the calibration vector [*Arhonditsis et al.*, 2007a]. The addition of a seasonally variant discrepancy term (Model 3) was more informative for some parameters, for example, the detritus mineralization (φ) and sinking rates (ψ), the phytoplankton maximum growth (a) and sinking rates (s), and half-saturation constant for PO_4 uptake

Table 4. Markov Chain Monte Carlo Posterior Estimates of the Mean Values and Standard Deviations of the Model Stochastic Nodes

	Prior		Model 1		Model 2		Model 3	
Parameter	Mean	SD	Mean	SD	Mean	SD	Mean	SD
a	1.226	0.256	0.890	0.082	1.179	0.197	1.018	0.108
d	0.172	0.018	0.160	0.014	0.171	0.016	0.175	0.016
pred	54.11	11.54	36.37	1.367	48.76	9.883	46.84	9.016
e	12.16	5.212	13.78	2.169	15.87	4.110	6.709	1.176
k	0.032	0.008	0.021	0.001	0.032	0.007	0.037	0.007
r	0.082	0.028	0.140	0.008	0.124	0.020	0.128	0.016
s	0.038	0.024	0.031	0.004	0.055	0.014	0.078	0.010
a	0.360	0.103	0.537	0.044	0.438	0.076	0.348	0.069
β	0.287	0.090	0.380	0.051	0.377	0.054	0.414	0.055
γ	0.287	0.090	0.392	0.065	0.405	0.063	0.437	0.079
λ	0.606	0.090	0.638	0.063	0.550	0.061	0.542	0.061
μ	6.500	1.540	9.425	0.463	6.574	1.054	6.474	1.306
φ	0.163	0.093	0.055	0.004	0.076	0.020	0.057	0.006
ψ	0.049	0.038	0.039	0.003	0.031	0.007	0.016	0.003
σ_{PO4}					2.059	0.600		
σ_{chla}					72.79	29.90		
σ_{zoop}					25.44	8.631		
σ_{det}					1.009	1.082		
ω_{PO4}							1.280	0.466
ω_{chla}							92.90	45.48
ω_{zoop}							18.80	6.845
ω_{det}							1.636	0.778
$PO4_{(0)}$	13.57	2.036	13.33	1.452	13.92	2.027	13.16	1.731
$PHYT_{(0)}$	61.94	9.291	85.23	13.23	68.35	12.68	63.63	12.23
$ZOOP_{(0)}$	15.97	2.396	16.73	1.437	18.74	3.174	19.97	6.368
$DET_{(0)}$	15.68	2.352	15.76	1.745	14.58	1.970	17.14	2.691

(e) (relative decrease >60%), while others remained unaltered with respect to their standard deviations. Generally, our results are fairly similar to those reported in the *Higdon et al.* [2004] study (see their Figures 3 and 11), suggesting that the effects of the discrepancy term on the parameter posteriors can be quite variant depending on the prior model specification and the system being modeled.

[31] In contrast with the GLUE analysis, it is interesting to note that the central tendency of several updated distributions, that is, phytoplankton respiration rate (r), detritus mineralization rate (φ), zooplankton excretion fraction (β), and regeneration of zooplankton predation excretion (γ) (Models 1, 2, and 3); zooplankton growth efficiency (a) and grazing half-saturation coefficient (μ) (Model 1), phytoplankton sinking loss rate (s) (Models 2 and 3), detritus sinking loss rate (ψ) and half-saturation constant for PO_4 uptake (e) (Model 3) were shifted relative to the prior assigned values. We also used the MCMC posterior samples from the Model 2 to determine the correlation structure of the model parameters (Table 5). The comparison with the results derived from the GLUE analysis (Table 3) shows that there was considerable consistency between the two correlation matrices, although there were relationships with stronger signals (e.g., zooplankton growth efficiency (a) and phytoplankton sinking loss rates (s)/maximum zooplankton grazing rate (λ)) and others that were not manifested under the MCMC sampling (e.g., detritus sinking rates (ψ) and detritus mineralization (φ)/zooplankton grazing half-saturation coefficient (μ)). Furthermore, we found similar to the parameters patterns regarding the CV values of the prior and posterior distributions of the initial conditions; the first model was more informative and resulted in reduced CV values (Figure 4 and Table 4).

[32] To gain insight into the third statistical formulation, we plotted the model estimates vis-à-vis the discrepancy (error) terms for the phosphate, detritus, phytoplankton, and zooplankton biomass concentrations (Figure 5). The model estimates (i.e., the term $f(\theta, x_i, y_0)$ in equation (16)) provided patterns similar to those found from the GLUE analysis; the model underestimated the observed spring phytoplankton ($\approx$300 μg C/L or 6 μg chla/L) and zooplankton ($\approx$70 μg C/L) dynamics. As hypothesized in the previous section, these results probably stem from the model error structure along with the inaccuracy of the boundary conditions and are partly accounted for by the discrepancy error terms (i.e., the term δ_{ij} in equation (16)); an indicator of how well the model is matching reality. For example, the model's inability to reproduce the Lake Washington spring plankton dynamics can explain the higher April–May δ_{phyt} and δ_{zoop} values, while the relatively higher $\delta_{April,PO4}$ also reflects the lower contemporaneous phosphate (PO_{4m}) concentrations. The posterior conditional variances (ω_j) of the seasonally variant discrepancies along with the four temporally constant error (σ_j) terms are shown in Figure 6. In the second statistical formulation, these error terms delineate a constant zone around the model estimates for the four state variables (i.e., the term $f(\theta, x_i, y_0)$ in equation (12)) that accounts for the discrepancy between the simulation model and the natural system, and it is worth noting the relatively high CV values of the detritus error term (σ_{det}).

[33] The three statistical formulations sampled with the MCMC scheme were favorably supported by the data and were accepted on the basis of their posterior predictive p values (0.103, 0.402 and 0.372 for Models 1, 2 and 3, respectively). The Bayes factor values B_{21} = 2.92, B_{31} = 2.47, B_{23} = 1.15 did not provide strong evidence in support of one of the three alternative models but did reflect a higher performance of the Model 2 [*Kass and Raftery*, 1995, p. 777]. The comparison between the observed (monthly averages over the 10-a period, 1994–2003) and posterior predictive monthly distributions for the three statistical formulations illustrates some features of the Bayesian calibration. The Model 1 provides relatively similar patterns to those found from the GLUE analysis, although the sampled prior distributions were less constrained (Figure 3). For example, regarding the central tendency of the predictive monthly distributions, we found that the predicted median values were almost consistently lower than the observed mean total phosphorus levels and also underestimated the spring maximum phytoplankton and zooplankton biomass. As noted before, these results can partly be attributed to the inaccurate representation of the boundary conditions (weather, exogenous loading, and vertical mixing) by simple periodic functions, while the relatively wide prediction bands for the spring plankton dynamics also reflect the higher observation error used for these months. The addition of the discrepancy terms in the second and third statistical formulation has significantly improved the results and now the model provides a good fit to the four state variables; all the observed monthly values were included within the 95% credible intervals (Figure 7). To more realistically account for the effects of the physical conditions on the Lake Washington patterns, we also employed a stochastic treatment of the forcing functions of the model (i.e., the trigonometric functions provided the

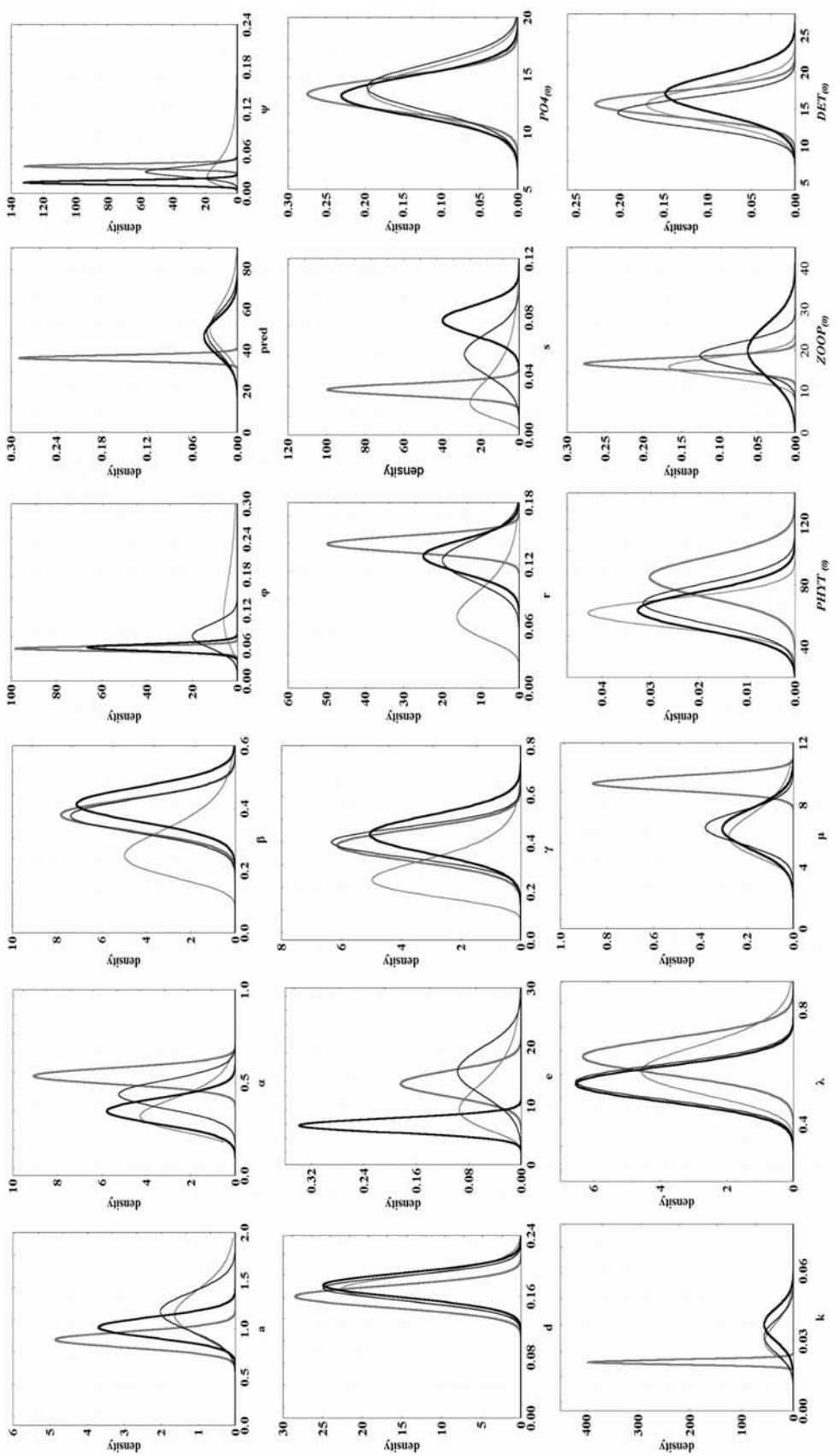

Figure 4. Prior (thin grey lines) and posterior (Model 1, thick grey lines; Model 2, thin black lines; and Model 3, thick black lines) distributions of the eutrophication model parameters and the initial conditions of the four state variables. The posteriors depict smoothed kernel density estimates based on 5000 Markov Chain Monte Carlo (MCMC) samples from each of the three models.

Table 5. Correlation Matrix of the Eutrophication Model Parameters Based on MCMC Posterior Samples, Model 2[a]

	a	d	Pred	e	k	r	s	α	β	γ	λ	μ	φ	Ψ
		−0.037	0.059	**0.583**	0.119	0.176	**0.398**	−0.003	0.120	0.005	−0.068	−0.021	−0.117	**−0.286**
a			0.044	−0.097	−0.058	0.046	0.079	0.158	−0.008	0.038	0.051	−0.212	−0.036	0.008
d				0.062	−0.134	0.011	−0.029	−0.155	0.037	−0.010	−0.095	**0.302**	0.006	−0.024
Pred					0.007	−0.108	**−0.351**	−0.223	−0.083	−0.055	−0.005	0.013	−0.090	0.018
e						0.017	0.177	0.039	0.125	−0.045	0.016	0.035	−0.093	0.030
k							−0.037	−0.045	0.063	−0.010	−0.111	0.087	−0.025	−0.067
s								**0.347**	0.208	0.096	−0.065	−0.056	0.015	**−0.397**
α									−0.085	−0.032	**−0.252**	**0.302**	0.081	0.180
β										0.031	−0.085	0.145	−0.123	0.057
γ											−0.117	0.027	−0.080	−0.057
λ												0.195	0.081	0.228
μ													**−0.247**	−0.050
φ														0.190
ψ														

[a]Bold numbers correspond to correlation coefficients with absolute value greater than 0.250. MCMC, Markov Chain Monte Carlo.

mean of a Gaussian distribution with standard deviation assumed to be 10% of the mean values). The predicted median spring plankton biomass levels were relatively closer to the lake seasonal dynamics but, not surprisingly, were also accompanied by wider prediction bands (not presented here).

4. Discussion

[34] Elucidation of the equifinality and uncertainty patterns in the multidimensional parameter spaces of mathematical models involves two critical decisions: (1) selection of the sampling scheme for generating input vectors, which then are evaluated with regards to the model performance, and (2) selection of the statistical problem description; that is, which likelihood function should we use and why? The former decision addresses the sampling efficiency of the approach (Random sampling, Latin hypercube, MCMC), while the latter one entails conceptual dilemmas (generalized or purely probabilistic likelihood functions) that can significantly alter the results. In this study, we examined the efficiency of two uncertainty analysis strategies with very distinct features, a generalized likelihood uncertainty estimation (GLUE) approach combined with a simple Monte Carlo sampling scheme and a Bayesian framework along with Markov Chain Monte Carlo (MCMC) simulations. Given the differences in the configuration of the two strategies, our aim was not to directly compare alternative model likelihood functions (or sampling schemes), but rather to illustrate the suite of techniques available for addressing equifinality and uncertainty in eutrophication models. Nonetheless, some of the (dis)similarities in the patterns of the posterior parameter distributions, correlation structure and predictive uncertainty can have prescriptive value and/or dictate future directions of the modeling practice.

4.1. Sampling Schemes

[35] Like other Bayesian-like algorithms, GLUE is typically combined with Monte Carlo sampling schemes, which draw samples uniformly and independently from the plausible parameter ranges. As a result, the Monte Carlo samples can misrepresent (or insufficiently cover) regions of high model likelihood; especially, when the joint prior parameter distribution is very wide or the parameters are highly correlated [*Qian et al.*, 2003]. In contrast with the large majority of the published GLUE applications, we attempted to optimize the efficiency of our sampling strategy by forming informative prior distributions from existing scientific knowledge, past experience and results from a preliminary exploratory analysis. With this configuration of our Monte Carlo sampling strategy, our intent was to focus on a subregion of the parameter space where there was evidence for higher likelihood of realistic reproduction of the system dynamics. Nonetheless, the resulting acceptance rate (≈1.3%) was still not higher than those reported in other studies, while the predictive outputs of the "behavioral" subset were characterized by two systematic errors; the misreproduction of the timing/magnitude of the spring plankton dynamics and the underestimation of the total phosphorus levels throughout the annual cycle. Given that the degree of updating of the model input parameters (i.e., reduction in the parameter uncertainty and shifts in the most likely value) was relatively limited, it was unclear whether these results were mainly driven by inefficient sampling of an ill-defined prior parameter space or by shortcomings of the eutrophication model structure and forcing functions.

[36] The latter issue was unequivocally clarified by the use of less constrained prior distributions along with the implementation of a MCMC sampling scheme; a method specifically designed to sample directly from the posterior distribution and to converge to the most probable region [*Gelman et al.*, 1995]. In particular, the statistical formulation that only considers the observation error (Model 1) provided relatively similar patterns and verified that the predictive bias is actually associated with the prior model specification (e.g., inadequacy of the model structure and/or inaccuracy of the boundary conditions). Generally, the MCMC procedure provided a convenient means to efficiently sample the parameter space of our intermediate complexity eutrophication model; we found that 30,000 samples with a fairly straightforward algorithm (i.e., general normal-proposal Metropolis) gave adequate summary statistics of the posterior parameter distributions and the predictive model outputs. Although more advanced procedures are available [*Gilks et al.*, 1998], several modeling studies from a variety of disciplines indicated that even simpler MCMC schemes can overcome the lack of analytical expressions for the posterior probability distribution [*Van Oijen et al.*, 2005]; typical problem with the

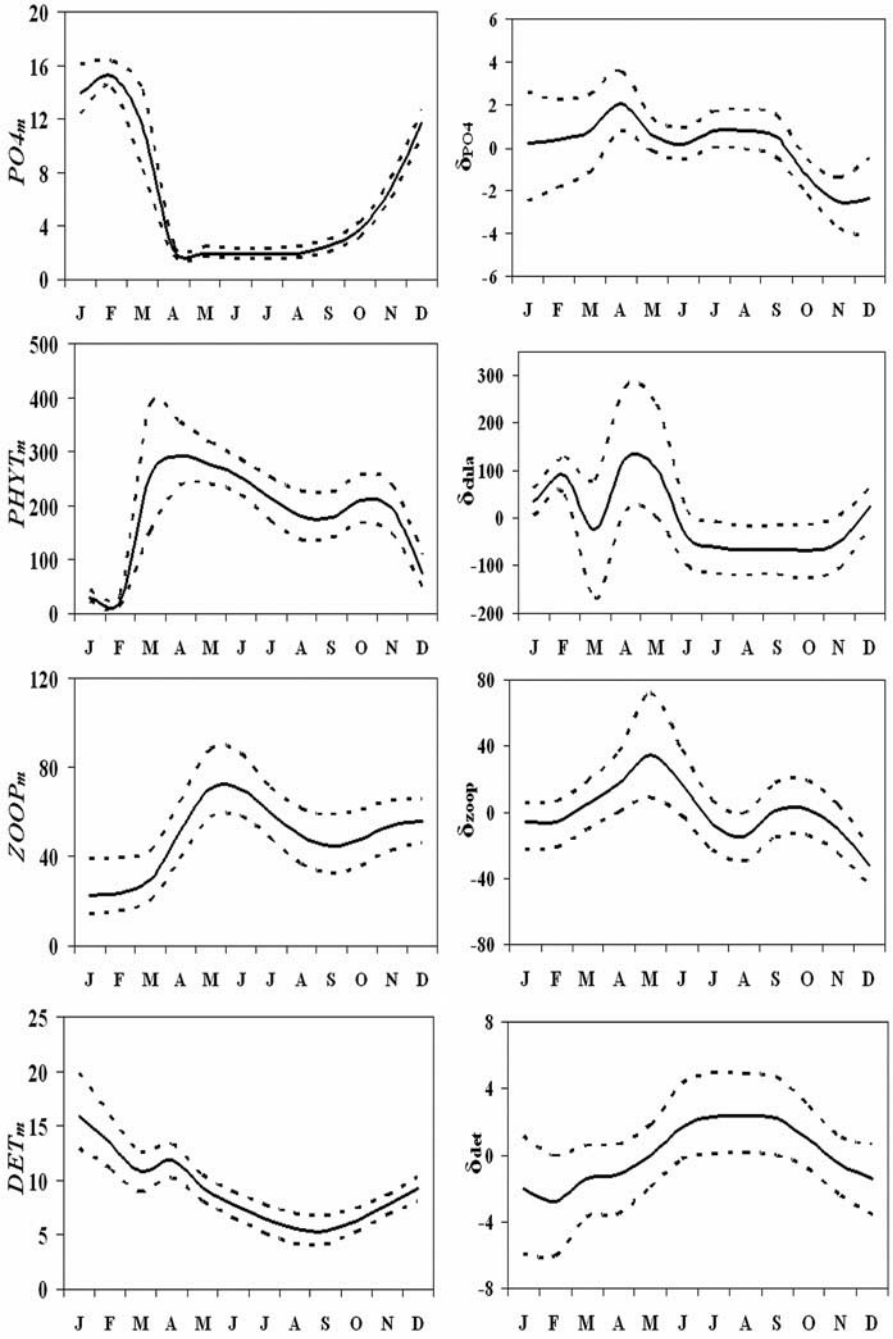

Figure 5. Time series plots of the model estimates (the term $f(\theta, x_i, y_0)$ in equation (16)) and the error terms (the term δ_{ij} in equation (16)), representing the discrepancy between the model and the natural system, for the phosphate, detritus, phytoplankton, and zooplankton biomass concentrations (Model 3).

nonlinear parameterizations used in eutrophication modeling. Moreover, other model inputs (e.g., initial conditions and boundary conditions) were also treated stochastically without the need to generate significantly more MCMC runs, which indicates that as long as the number of parameters that drive the model outputs does not change significantly, then the number of runs required to sufficiently approximate the posterior will be roughly the same [*Jansen and Hagenaars*, 2004]. In contrast, the simple Monte Carlo sampling followed in the GLUE presentation was less efficient; characteristic example was the uninformative patterns of the univariate marginal parameter distributions (see Figure 2), despite the use of 60,000 samples (twice the number of the MCMC samples) from relatively truncated prior parameter distributions and the compromises made regarding the way initial conditions and boundary conditions were handled. Although beyond the scope of the present paper, an interesting topic for future research will be the combination of MCMC schemes with the GLUE methodology [*Vrugt et al.*, 2002].

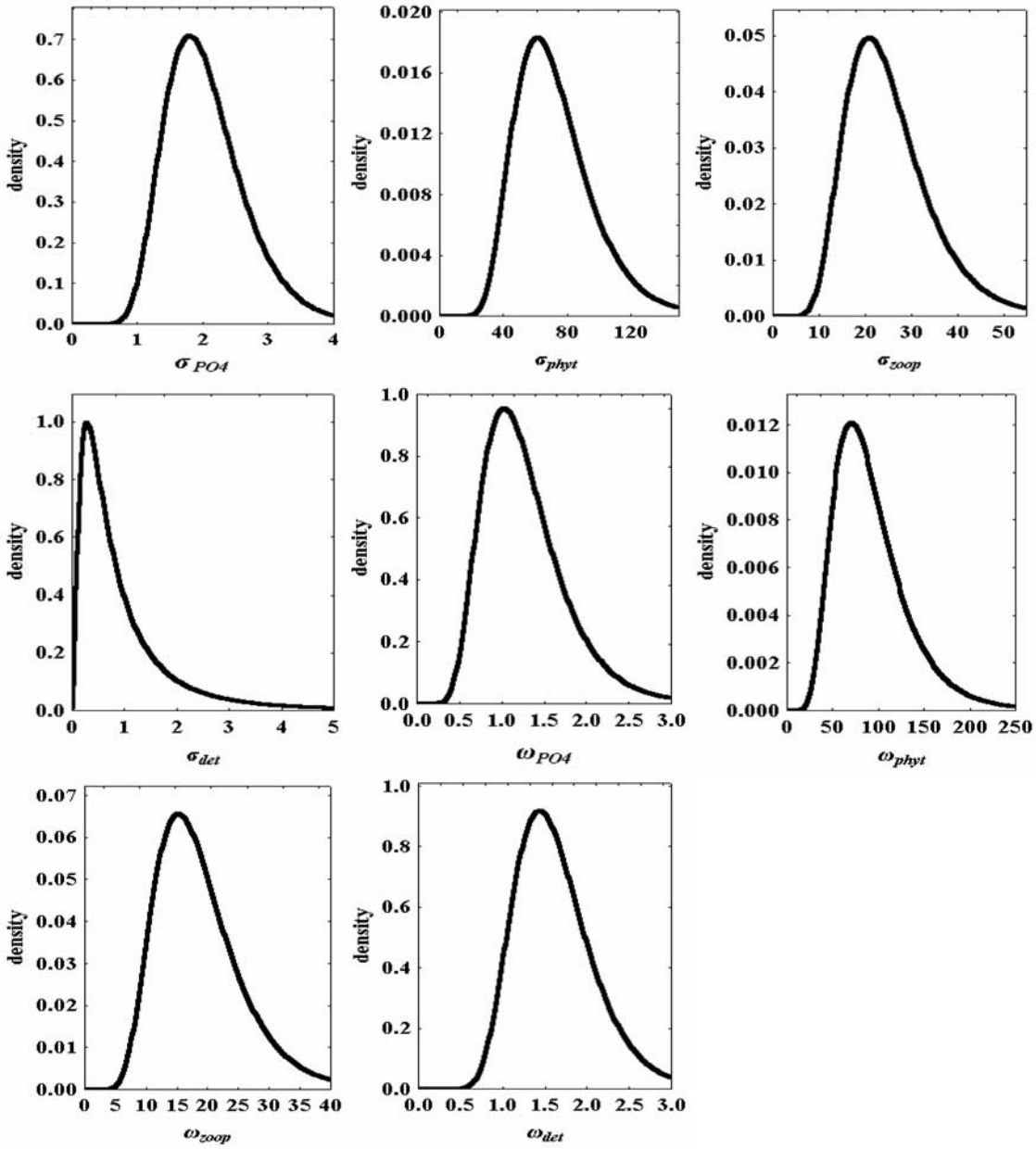

Figure 6. Posterior distributions of the σ (seasonally invariant discrepancy between the simulation model and the natural system) and ω (conditional variance of the seasonally variant discrepancy) terms of the second and third statistical formulations, respectively.

4.2. Likelihood Functions

[37] GLUE is a conceptually straightforward methodology in which a broad range of likelihood measures can be used to define the model error structure and to delineate plausible regions of parameter values [*Beven*, 2001]. Generally, it is believed that the GLUE methodology is perhaps conceptually better suited to highly dimensional input spaces with significant equifinality problems, because model response surfaces are more complex than smooth, single-peaked distributions (see discussion in the paper by *Kennedy and O'Hagan* [2001]). From a statistical inference viewpoint, however, the lack of formal representation of the model error has been criticized for providing biased parameter estimates when not taking into account the correct model error structure and for resulting in likelihoods that do not correspond to the true probabilities of predicting an output given the model [*Thiemann et al.*, 2001]. Some of the effects of the nonprobabilistic likelihood functions were also manifested in our study. For example, the likelihood function used for the presented results (equation (1)) depends heavily on an expectation that good model solutions do not exist; thus, if a good parameter set is actually found then the assigned likelihood weight can be very high relative to all other sets. This effect can already be seen in the best simulations of the scatter plots (Figure 2). Furthermore, the use of alternative likelihood measures (i.e., the relative error and the modeling efficiency) resulted in a 60–80% overlap among the three subsets of behavioral runs derived from the different measures of fit, despite the

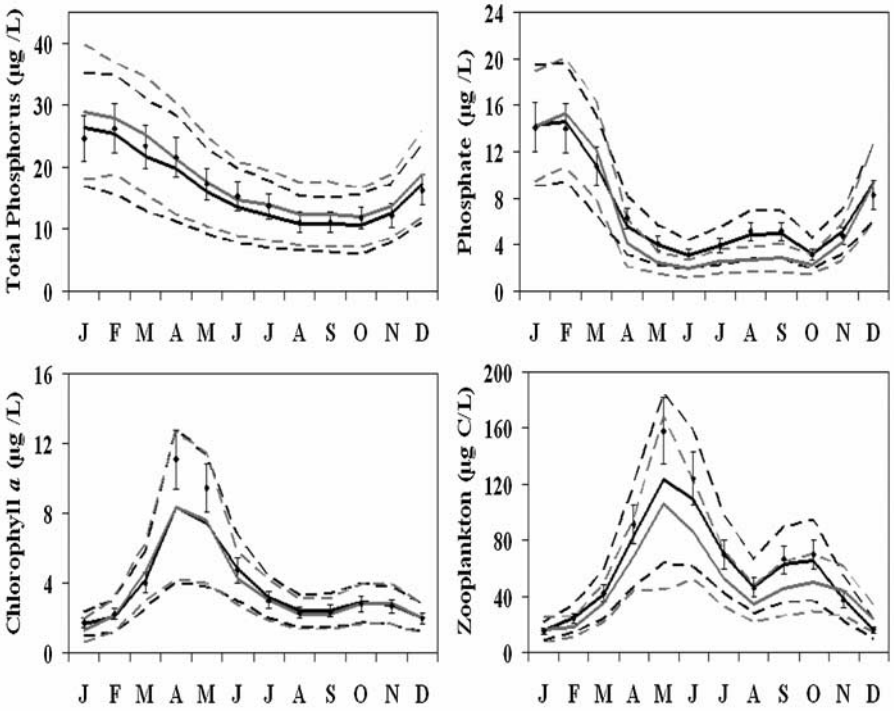

Figure 7. Comparison between the observed and posterior predictive monthly distributions for phosphate, total phosphorus, chlorophyll *a*, and zooplankton biomass based on 5000 Markov Chain Monte Carlo posterior samples from Model 2 (black line) and Model 3 (grey line). Single dots and the respective black lines correspond to the monthly averages and standard deviations, reflecting the analytical error and the interannual variability in the lake over the 10-a period, 1994–2003.

relatively similar posterior parameter statistics and predictive uncertainty bounds.

[38] On the other hand, the appropriateness of the formal (e.g., Gaussian) likelihoods for complex overparameterized mechanistic models frequently used in eutrophication research has not been explored yet in the modeling literature. In this study, we started with a statistical formulation that assumed a ''perfect'' model structure and additive (or multiplicative) measurement errors. Combined with a MCMC scheme, this approach generates a series of model realizations each of which is evaluated, assuming that the model is correct [*Beven*, 2006]. The stretching of the model likelihood surface implied by these assumptions should lead to overconditioning of the parameter estimates owing to an overestimation of the information content of the observations/residuals given the potential for model error. The latter assertion can probably explain the narrow-shaped posterior parameter distributions derived from the first statistical formulation (Figure 4), which then may lead to nonbehavioral simulations when used under different conditions.

[39] The introduction of two models that explicitly acknowledge the lack of perfect simulators of natural system dynamics is a conceptual advancement over the first statistical formulation. In this study, we found that the inclusion of error terms that explicitly account for the (variant or invariant with the input conditions) discrepancy between mathematical model and environmental system improved the model predictions and all the observed monthly values were included within the 95% credible intervals. However, it should be emphasized that despite the promising results, this modeling framework will possibly require substantial modifications to accommodate highly multivariate outputs. Although, in a follow up study, we found that the inclusion of several more state variables does not have an effect on the consistency of the results [*Arhonditsis et al.*, 2007b], the consideration of multiple error sources often entails overparameterized formulations, for example, the third model in which 12 parameters per state variable are used to describe the seasonal variability in the error variance. In this direction, several interesting ideas have been proposed in the literature, such as dimension reduction strategies, adaptive designs to overcome limited number of simulation runs, and replacement of the simulators with statistical models that encompass key features of the modeled system [*Craig et al.*, 2001; *Wikle et al.*, 2001; *Goldstein and Rougier*, 2004; *Higdon et al.*, 2004]. In the modeling practice, our experience is that only a subset of the input parameters is influential on the model outputs [*Arhonditsis and Brett*, 2005a], and therefore an effective calibration does not necessarily require statistical formulations framed in a hyperdimensional context [*Kennedy and O'Hagan*, 2001]. An alternative strategy will resemble the modeling framework used in this study; intermediate complexity models that provide realistic platforms for reproducing real world dynamics while conforming to the parsimony principle.

[40] In conclusion, we examined the efficiency of two methodological frameworks, a generalized likelihood uncertainty estimation (GLUE) approach combined with a simple Monte Carlo sampling scheme and a Bayesian

methodological framework along with Markov Chain Monte Carlo (MCMC) simulations, to quantify the information the data contain about model inputs, to offer insights into the covariance structure among parameter estimates, and to obtain predictions along with uncertainty bounds for modeled output variables. GLUE has been criticized for its statistical correctness owing to the use of a variety of goodness-of-fit measures, but the same relaxation of the likelihood functions used make this methodology appropriate for a wide range of model complexity. On the other hand, the formal probabilistic models provide sound statistical inference, but the stretching of the model response surfaces by inappropriate likelihoods can provide misleading results and undermine their credibility for predicting future conditions. The inclusion of multiple error sources (e.g., measurement error, parametric error, and model structure imperfection) into the probabilistic likelihoods provides a promising framework for assessing predictive uncertainty, and future research should involve the examination of its suitability for more complex models extensively used in eutrophication management. Our study also highlights the efficiency of MCMC sampling schemes, specifically designed to sample directly from the posterior distribution, to fully employ Bayesian inference techniques which then can be easily integrated with, at least, intermediate complexity mechanistic models ($\leq$10 state variables). Given the substantial model forecast uncertainty in most water quality models, the arbitrary selection of higher, and often unattainable, threshold values for environmental variables (quality goals/standards), risky model-based management decisions and unanticipated system responses are the norm in current management practice. The development of novel methods that can accommodate rigorous and complete error analysis is an imperative challenge for the future of environmental modeling [*Pappenberger and Beven*, 2006; *Arhonditsis et al.*, 2006].

Appendix A: Specific Functional Forms of the Eutrophication Model

[41] The specific functional forms of the eutrophication model.

$$\frac{dPO_4}{dt} = -\frac{PO_4}{e+PO_4}\mathrm{a}\sigma_{(t)}PHYT\ P/C_{phyto} + \frac{\beta\lambda\left(\left(PHYT\cdot P/C_{phyto}\right)^2+\omega DET^2\right)}{\mu^2+\left(PHYT\cdot P/C_{phyto}\right)^2+\omega DET^2}\sigma_{(tz)}ZOOP\ P/C_{zoop} \quad (A1)$$

$$+\gamma d\sigma_{(tz)}\frac{ZOOP^3}{pred^2+ZOOP^2}P/C_{zoop} + \phi\sigma_{(t)}DET + k\left(1-\sigma_{(t)}\right)\left(PO_{4(hypo)}-PO_4\right)+PO_{4exog} \quad (A2)$$

$$-outflows\cdot PO_4 \quad (A3)$$

$$\sigma_{(t)} = \frac{\left(1-\varepsilon\cos\left(\frac{2\pi t}{365}\right)\right)}{1+\varepsilon}\quad \sigma_{(tz)} = \frac{\left(1-\varepsilon\cos\left(\frac{2\pi t}{365}-0.5\right)\right)}{1+\varepsilon} \quad (A4)$$

$$\begin{aligned} PO_{4(hypo)} &= 11+3\sin(2\pi(t/365+0.3)) \\ PO_{4exog} &= 0.047+0.02\sin(2\pi(t/365+0.12)) \\ outflows &= 0.0028+0.0014\sin(2\pi(t/365+0.12)) \end{aligned} \quad (A5)$$

$$\frac{dPHYT}{dt} = \frac{PO_4}{e+PO_4}\mathrm{a}\sigma_{(t)}PHYT - r\sigma_{(t)}PHYT - \frac{\lambda\left(PHYT\cdot P/C_{phyto}\right)^2}{\mu^2+\left(PHYT\cdot P/C_{phyto}\right)^2+\omega DET^2}\sigma_{(tz)}ZOOP \quad (A6)$$

$$-sPHYT - outflows\cdot PHYT \quad (A7)$$

$$\frac{dZOOP}{dt} = \frac{\alpha\lambda\left(\left(PHYT\cdot P/C_{phyto}\right)^2+\omega DET^2\right)}{\mu^2+\left(PHYT\cdot P/C_{phyto}\right)^2+\omega DET^2}\sigma_{(tz)}ZOOP - d\sigma_{(tz)}\frac{ZOOP^3}{pred^2+ZOOP^2} - outflows\cdot ZOOP \quad (A8)$$

$$\frac{dDET}{dt} = r\sigma_{(t)}PHYT\ P/C_{phyto} + \frac{\left[(1-\alpha-\beta)\left(PHYT\cdot P/C_{phyto}\right)^2-(\alpha+\beta)\omega DET^2\right]\lambda}{\mu^2+\left(PHYT\cdot P/C_{phyto}\right)^2+\omega DET^2}\cdot\sigma_{(tz)}ZOOP\ P/C_{zoop} \quad (A9)$$

$$-\varphi\sigma_{(t)}DET - \psi DET + DET_{exog} - outflows\cdot DET \quad (A10)$$

$$DET_{exog} = 0.2+0.12\sin(2\pi(t/365+0.16)) \quad (A11)$$

[42] **Acknowledgments.** Funding for this study was provided by the National Sciences and Engineering Research Council of Canada (NSERC, Discovery Grants), the Connaught Committee (University of Toronto, Matching Grants 2006–2007), and the Summer Career Placement Program (Human Resources Development Canada). Keith Beven and two anonymous reviewers offered excellent comments on an earlier version of the manuscript. We also wish to thank Maria Kouzelli-Arhonditsi for interesting suggestions on the presentation of our results. All the material pertinent to this study is available upon request from the first author.

References

Arhonditsis, G. B., and M. T. Brett (2004), Evaluation of the current state of mechanistic aquatic biogeochemical modeling, *Mar. Ecol. Prog. Ser.*, *271*, 13–26.

Arhonditsis, G. B., and M. T. Brett (2005a), Eutrophication model for Lake Washington (USA): part I, Model description and sensitivity analysis, *Ecol. Modell.*, *187*, 140–178.

Arhonditsis, G. B., and M. T. Brett (2005b), Eutrophication model for Lake Washington (USA): part II, Model calibration and system dynamics analysis, *Ecol. Modell.*, *187*, 179–200.

Arhonditsis, G., M. T. Brett, and J. Frodge (2003), Environmental control and limnological impacts of a large recurrent spring bloom in Lake Washington, USA, *Environ. Manage.*, *31*, 603–618.

Arhonditsis, G. B., B. A. Adams-VanHarn, L. Nielsen, C. A. Stow, and K. H. Reckhow (2006), Evaluation of the current state of mechanistic aquatic biogeochemical modeling: citation analysis and future perspectives, *Environ. Sci. Technol.*, *40*, 6547–6554.

Arhonditsis, G. B., S. S. Qian, C. A. Stow, E. C. Lamon, and K. H. Reckhow (2007a), Eutrophication risk assessment using Bayesian cali-

bration of process-based models: Application to a mesotrophic lake, *Ecol. Modell.*, *208*, 215–229.

Arhonditsis, G. B., D. Papantou, W. Zhang, G. Perhar, E. Massos, and M. Shi (2007b), Bayesian calibration of mechanistic aquatic biogeochemical models and benefits for environmental management, *J. Mar. Syst.*, in press.

Beck, M. B. (1987), Water-quality modeling: A review of the analysis of uncertainty, *Water Resour. Res.*, *23*, 1393–1442.

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

Beven, K. J. (2001), *Rainfall-Runoff Modeling: The Primer*, 360 pp., John Wiley, New York.

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

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

Brett, M. T., S. E. Mueller, and G. B. Arhonditsis (2005), A daily time series analysis of stream water phosphorus concentrations along an urban to forest gradient, *Environ. Manage.*, *35*, 56–71.

Brooks, S. P., and A. Gelman (1998), General methods for monitoring convergence of iterative simulations, *J. Comput. Graph Stat.*, *7*, 434–455.

Brun, R., P. Reichert, and H. R. Kunsch (2001), Practical identifiability analysis of large environmental simulation models, *Water Resour. Res.*, *37*, 1015–1030.

Casella, G., and E. I. George (1992), Explaining the Gibbs sampler, *Am. Stat.*, *46*, 167–174.

Chapra, S. C. and R. P. Canale (1998), *Numerical Methods for Engineers*, 3rd ed., 924 pp., McGraw-Hill, New York.

Craig, P. S., M. Goldstein, J. C. Rougier, and A. H. Seheult (2001), Bayesian forecasting for complex systems using computer simulators, *J. Am. Stat. Assoc.*, *96*, 717–729.

Dilks, D. W., R. P. Canale, and P. G. Meier (1992), Development of Bayesian Monte-Carlo techniques for water-quality model uncertainty, *Ecol. Modell.*, *62*, 149–162.

Di Toro, D. M. and G. van Straten (1979), Uncertainty in the parameters and predictions of phytoplankton models, *WP-79-27*, Int. Inst. for Appl. Syst. Anal., Laxenburg, Austria.

Edmondson, W. T. (1994), Sixty years of Lake Washington: A curriculum vitae, *Lake Reservoir Manage.*, *10*, 75–84.

Edwards, A. M., and A. Yool (2000), The role of higher predation in plankton population models, *J. Plankton Res.*, *22*, 1085–1112.

Engeland, K., and L. Gottschalk (2002), Bayesian estimation of parameters in a regional hydrological model, *Hydrol. Earth Syst. Sci.*, *6*, 883–898.

Franks, S. W., P. Gineste, K. J. Beven, and P. Merot (1998), On constraining the predictions of a distributed models: The incorporation of fuzzy estimates of saturated areas into the calibration process, *Water Resour. Res.*, *34*, 787–797.

Freer, J., J. McDonnell, K. J. Beven, D. Brammer, D. Burns, R. P. Hooper, and C. Kendal (1997), Topographic controls on subsurface storm flow at the hillslope scale for two hydrologically distinct small catchments, *Hydrol. Processes*, *11*, 1347–1352.

Gelfand, A. E., and A. F. M. Smith (1990), Sampling based approaches to calculating marginal densities, *J. Am. Stat. Assoc.*, *85*, 398–409.

Gelman, A. (2005), Prior distributions for variance parameters in hierarchical models, *Bayesian Anal.*, *1*, 1–19.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (1995), *Bayesian Data Analysis*, 518 pp., Chapman and Hall, New York.

Gelman, A., X. L. Meng, and H. Stern (1996), Posterior predictive assessment of model fitness via realized discrepancies, *Stat. Sin.*, *6*, 733–807.

Gilks, W. R., S. Richardson, and D. J. Spiegelhalter (1998), *Markov Chain Monte Carlo in Practice*, 512 pp., CRC, Boca Raton, Fla.

Goldstein, M., and J. Rougier (2004), Probabilistic formulations for transferring inferences from mathematical models to physical systems, *SIAM J. Sci. Comput.*, *26*, 467–487.

Grover, J. P. (1991), Resource competition in a variable environment: Phytoplankton growing according to the Variable-Internal-Stores model, *Am. Nat.*, *138*, 811–835.

Hastings, W. K. (1970), Monte-Carlo Sampling methods using Markov Chains and their applications, *Biometrika*, *57*, 97–109.

Higdon, D., M. Kennedy, J. C. Cavendish, J. A. Cafeo, and R. D. Ryne (2004), Combining field data and computer simulations for calibration and prediction, *SIAM J. Sci. Comput.*, *26*, 448–466.

Hong, B. G., R. L. Strawderman, D. P. Swaney, and D. A. Weinstein (2005), Bayesian estimation of input parameters of a nitrogen cycle model applied to a forested reference watershed, Hubbard Brook Watershed Six, *Water Resour. Res.*, *41*, W03007, doi:10.1029/2004WR003551.

Hornberger, G. M., and R. C. Spear (1981), An approach to the preliminary analysis of environmental systems, *J. Environ. Manage.*, *12*, 7–18.

Jansen, M. J. W., and T. J. Hagenaars (2004), Calibration in a Bayesian modeling framework, in *Bayesian Statistics and Quality Modeling in the Agro-Food Production Chain*, edited by M. A. J. S. Van Boekel, A. Stein, and A. H. C. Van Bruggen, pp. 47–55, Springer, New York.

Jorgensen, S. E., and G. Bendoricchio (2001), *Fundamentals of Ecological Modelling*, 3rd ed., 530 pp., Elsevier, New York.

Kass, R. E., and A. E. Raftery (1995), Bayes factors, *J. Am. Stat. Assoc.*, *90*, 773–795.

Kennedy, C. M., and A. O'Hagan (2001), Bayesian calibration of computer models, *J. R. Stat. Soc., Ser. B*, *63*, 425–464.

Lamb, R., K. Beven, and S. Myrabo (1998), Use of spatially distributed water table observations to constrain uncertainty in a rainfall-runoff model, *Adv. Water Resour.*, *22*, 305–317.

Meinhold, R. J., and N. D. Singpurwalla (1983), Understanding the Kalman Filter, *Am. Stat.*, *37*, 123–127.

Metropolis, N. (1953), Equation of state calculations by fast computing machines, *J. Chem. Phys.*, *21*, 1087–1092.

Neal, R. (1998), Suppressing random walks in Markov chain Monte Carlo using ordered over-relaxation, in *Learning in Graphical Models*, edited by M. I. Jordan, pp. 205–230, Springer, New York.

Omlin, M., and P. Reichert (1999), A comparison of techniques for the estimation of model prediction uncertainty, *Ecol. Modell.*, *115*, 45–59.

Page, T., K. J. Beven, and J. D. Whyatt (2004), Predictive capability in estimating changes in water quality: Long-term responses to atmospheric deposition, *Water Air Soil Pollut.*, *151*, 215–244.

Pappenberger, F., and K. J. Beven (2006), Ignorance is bliss: Or seven reasons not to use uncertainty analysis, *Water Resour. Res.*, *42*, W05302, doi:10.1029/2005WR004820.

Pole, A., M. West, and P. J. Harrison (1994), *Applied Bayesian Forecasting and Time Series Analysis*, 409 pp., Chapman and Hall, New York.

Qian, S. S., C. A. Stow, and M. E. Borsuk (2003), On Monte Carlo methods for Bayesian inference, *Ecol. Modell.*, *159*, 269–277.

Ratto, M., S. Tarantola, and A. Saltelli (2001), Sensitivity analysis in model calibration: GSA-GLUE approach, *Comput. Phys. Commun.*, *136*, 212–224.

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

Reichert, P., and M. Omlin (1997), On the usefulness of overparameterized ecological models, *Ecol. Modell.*, *95*, 289–299.

Reichert, P., M. Schervish, and M. J. Small (2002), An efficient sampling technique for Bayesian inference with computationally demanding models, *Technometrics*, *44*, 318–327.

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

Shaddick, G., and J. Wakefield (2002), Modelling daily multivariate pollutant data at multiple sites, *J. R. Stat. Soc., Ser. C*, *51*, 351–372.

Sorooshian, S., and V. K. Gupta (1995), Model calibration, in *Computer Models of Watershed Hydrology*, edited by V. P. Singh, pp. 23–68, Water Resources Publications, Highlands Ranch, Colo.

Spear, R. C. (1997), Large simulation models: Calibration, uniqueness and goodness of fit, *Environ. Modell. Software*, *12*, 219–228.

Spear, R. C., and G. M. Hornberger (1980), Eutrophication in Peen Inlet: 2. Identification of critical uncertainties via generalized sensitivity analysis, *Water Res.*, *14*, 43–49.

Spear, R. C., T. M. Grieb, and N. Shang (1994), Parameter uncertainty and interaction in complex environmental-models, *Water Resour. Res.*, *30*, 3159–3169.

Steinberg, L. J., K. H. Reckhow, and R. L. Wolpert (1996), Bayesian model for fate and transport of polychlorinated biphenyl in upper Hudson River, *J. Environ. Eng.*, *122*, 341–349.

Steinberg, L. J., K. H. Reckhow, and R. L. Wolpert (1997), Characterization of parameters in mechanistic models: A case study of a PCB fate and transport model, *Ecol. Modell.*, *97*, 35–46.

Stow, C. A., C. Roessler, M. E. Borsuk, J. D. Bowen, and K. H. Reckhow (2003), Comparison of estuarine water quality models for total maximum daily load development in Neuse River Estuary, *J. Water Resour. Plann. Manage.*, *129*, 307–314.

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

Van Oijen, M., J. Rougier, and R. Smith (2005), Bayesian calibration of process-based forest models: Bridging the gap between models and data, *Tree Physiol.*, *25*, 915–927.

Vrugt, J. A., W. Bouten, H. V. Gupta, and S. Sorooshian (2002), Toward improved identifiability of hydrologic model parameters: The information content of experimental data, *Water Resour. Res.*, *38*(12), 1312, doi:10.1029/2001WR001118.

Wikle, C. K., R. F. Milliff, D. Nychka, and L. M. Berliner (2001), Spatiotemporal hierarchical Bayesian modeling: Tropical ocean surface winds, *J. Am. Stat. Assoc.*, *96*, 382–397.

Zak, S. K., and K. J. Beven (1999), Equifinality, sensitivity and predictive uncertainty in the estimation of critical loads, *Sci. Total Environ.*, *236*, 191–214.

Zak, S. K., K. Beven, and B. Reynolds (1997), Uncertainty in the estimation of critical loads: A practical methodology, *Water Air Soil Pollut.*, *98*, 297–316.

G. B. Arhonditsis, A. Das, E. Massos, G. Perhar, M. Shi, and W. Zhang, Department of Physical and Environmental Sciences, University of Toronto, 1265 Military Trail, Toronto, ON, Canada M1C 1A4. (georgea@utsc.utoronto.ca)