J. Great Lakes Res. 34:698–720
Internat. Assoc. Great Lakes Res., 2008

Predicting the Frequency of Water Quality Standard Violations Using Bayesian Calibration of Eutrophication Models

Weitao Zhang[1] and George B. Arhonditsis[1, 2,*]

[1]*Department of Geography*
University of Toronto
Toronto, Ontario M5S 3G3

[2]*Department of Physical & Environmental Sciences*
University of Toronto
Toronto, Ontario M1C 1A4

ABSTRACT. *The water quality standard setting process usually relies on mathematical models with strong mechanistic basis, as this provides assurance that the model will more realistically project the effects of alternative management schemes. From an operational standpoint, the interpretation of model results should be coupled with rigorous error analysis and explicit consideration of the predictive uncertainty and natural variability. In this study, our main objective is to attain effective model calibration and rigorous uncertainty assessment by integrating environmental mathematical modeling with Bayesian analysis. We use a complex aquatic biogeochemical model that simulates multiple elemental cycles (org. C, N, P, Si, O), multiple functional phytoplankton (diatoms, green algae and cyanobacteria) and zooplankton (copepods and cladocerans) groups. The Bayesian calibration framework is illustrated using three synthetic datasets that represent oligo-, meso- and eutrophic lake conditions. Scientific knowledge, expert judgment, and observational data were used to formulate prior probability distributions and characterize the uncertainty pertaining to a subset of the model parameters, i.e., a vector comprising the 35 most influential parameters based on an earlier sensitivity analysis of the model. Our study also underscores the lack of perfect simulators of natural system dynamics using a statistical formulation that explicitly accounts for the discrepancy between mathematical models and environmental systems. The model reproduces the key epilimnetic temporal patterns and provides realistic estimates of predictive uncertainty for water quality variables of environmental management interest. Our analysis also demonstrates how the Bayesian parameter estimation can be used for assessing the exceedance frequency and confidence of compliance of different water quality criteria. The proposed methodological framework can be very useful in the policy-making process and can facilitate environmental management decisions in the Laurentian Great Lakes region.*

INDEX WORDS: *Environmental management, process-based models, eutrophication, Bayesian calibration, uncertainty analysis, Markov Chain Monte Carlo, water quality standards.*

INTRODUCTION

In his 2006 review paper, D.W. Schindler highlighted the cultural eutrophication as one of the preeminent threats to the integrity of freshwater ecosystems worldwide. He also emphatically argued that our current understanding and management of eutrophication has evolved from simple control of point and non-point nutrient sources to the explicit recognition that it often stems from the cumulative effects of the human activities on climate, global element cycles, land use, and fisheries. Therefore, alleviating eutrophication problems often involves complex policy decisions aiming to protect the functional properties of the freshwater ecosystem community as well as to restore many of the features of the surrounding watershed. In the Great Lakes region, the growing appreciation of the complexity pertaining to eutrophication control and the need for addressing the combined effects of a suite of tightly intertwined stressors has sparked considerable confusion and disagreements (Hartig *et al.* 1998, Bowerman *et al.* 1999). Much of this

*Corresponding author. E-mail: georgea@utsc.utoronto.ca

controversy has arisen as to whether the Great Lakes Water Quality Agreement is a thrust for improving water quality or for maintaining ecosystem integrity, and the proposed transition from the Water Quality/Fisheries Exploitation paradigms into the Ecosystem Management paradigm has been repeatedly debated in the literature (Bowerman *et al*. 1999, Minns and Kelso 2000). The defenders of the traditional paradigms have argued that the shift of focus from water quality to ecosystem management has also been accompanied by a shift from the traditional identification of simple cause–effect relationships to a multi-causal way of thinking to accommodate the complexity of ecosystems. In this context, the crux of the problem is that the ecological complexity along with the underlying uncertainty can be a major impediment for deriving the straightforward scientific answers required from the regulatory agencies to implement the provisions of the Great Lakes Water Quality Agreement (Bowerman *et al.* 1999, Krantzberg 2004).

Aside from the environmental thinking, the emergence of the ecosystem approach has also pervaded the contemporary mathematical modeling practice, increasing the demand for more complex ecosystem models. Earlier eutrophication modeling studies in the Great Lakes provided long-term forecasts and insightful retrospective analysis using as foundation the interplay among nutrient loading, hydrodynamics, phytoplankton response, and sediment oxygen demand (Bierman and Dolan 1986, Lam *et al.* 1987a, DiToro *et al.* 1987). Yet, the current challenges make compelling the development of more realistic platforms (i) to elucidate causal mechanisms, complex interrelationships, direct and indirect ecological paths of the Great Lakes basin ecosystem; (ii) to examine the interactions among the various stressors (e.g., climate change, urbanization/land-use changes, alternative management practices, invasion of exotic species); and (iii) to assess their potential consequences on the lake ecosystem functioning (e.g., food web dynamics, benthic-pelagic coupling, fish communities) (Mills *et al.* 2003, Leon *et al.* 2005). In this regard, a characteristic example is the integrated eutrophication-zebra mussel bioenergetic model developed for identifying the factors that promote the re-occurrence of *Microcystis* blooms in the Saginaw Bay, Lake Huron (Bierman *et al.* 2005). It was shown that the zebra mussels through selective cyanobacteria rejection, increased sediment-water phosphorus fluxes can cause structural shifts in the phytoplankton community, and the impact of these perturbations varies depending on the magnitude of the zebra mussel densities and their distribution among different age groups. The Bierman *et al.* (2005) study is an example of how the increase of the articulation level of our mathematical models allows performing experiments that are technologically or economically unattainable by other means, thereby gaining insights into the direct and synergistic effects induced from the multitude of stressors on the various lake ecosystem components.

While the development of more holistic modeling constructs is certainly the way forward, the question arising is: do we have the knowledge to parameterize or even to mathematically depict the new biotic relationships and their interactions with the abiotic environment? More importantly, how reliable are the long-term projections generated from the current generation of mathematical models? Our experience is that the performance of existing mechanistic biogeochemical models declines as we move from physical-chemical to biological components of aquatic ecosystems (Arhonditsis and Brett 2004). Because of the still poorly understood ecology, we do not have robust parameterizations to support predictions in a wide range of spatiotemporal domains (Anderson 2005). Despite the repeated efforts to explicitly treat multiple biogeochemical cycles, to increase the functional diversity of biotic communities, and to refine the mathematical description of the higher trophic levels, modelers still haven't gone beyond the phase of identifying the unforeseeable ramifications and the challenges that we need to confront so as to strengthen model foundation (Anderson 2006). Furthermore, the additional model complexity will increase the disparity between what ideally we want to learn (internal description of the system and model endpoints) and what can realistically be observed, thereby reducing our ability to properly constrain the model parameters from observations (Denman 2003). The poor model identifiability undermines the predictive power of our models and their ability to support environmental management decisions (Arhonditsis *et al.* 2006). Thus, the most prudent strategy is to incorporate complexity gradually and this process should be accompanied by critical evaluation of the model outputs; the latter concern highlights the central role of uncertainty analysis.

Uncertainty analysis of mathematical models has received considerable attention in aquatic ecosystem research, and there have been several attempts to rigorously address issues pertaining to model

structure and input error (Beck 1987, Reichert and Omlin 1997, Stow *et al.* 2007). In this direction, Arhonditsis *et al.* (2007) recently introduced a Bayesian calibration scheme using intermediate complexity mathematical models (4-8 state variables) and statistical formulations that explicitly accommodate measurement error, parameter uncertainty, and model structure imperfection. The Bayesian calibration methodology offers several technical advances, such as alleviation of the identification problem, sequential updating of the models, realistic uncertainty estimates of ecological predictions, and ability to obtain weighted averages of the forecasts from different models, that can be particularly useful for environmental management (Arhonditsis *et al.* 2007; 2008a, b). Nonetheless, the capacity of this approach to be coupled with complex mathematical models has not been demonstrated yet and recent studies have cautioned that this modeling framework will possibly require substantial modifications to accommodate highly multivariate outputs (Arhonditsis *et al.* 2008b).

In this paper, our main objective is to integrate the Bayesian calibration framework with a complex aquatic biogeochemical model that simulates multiple elemental cycles (org. C, N, P, Si, O), multiple functional phytoplankton (diatoms, green algae, and cyanobacteria) and zooplankton (copepods and cladocerans) groups. Because the model structure and complexity is suitable for addressing a variety of eutrophication-related problems (chlorophyll *a*, water transparency, cyanobacteria dominance, hypoxia), our presentation is highly relevant to the Great Lakes modeling practice. This illustration is based on three synthetic datasets representing oligo-, meso-, and eutrophic lake conditions. Our analysis also shows how the Bayesian parameter estimation can be used for assessing the exceedance frequency and confidence of compliance of different water quality criteria. We conclude by pinpointing some of the anticipated benefits from the proposed approach, such as the assessment of uncertainty in model predictions and expression of model outputs as probability distributions, the optimization of the sampling design of monitoring programs, and the alignment with the policy practice of adaptive management, which can be particularly useful for stakeholders and policy makers when making decisions for sustainable environmental management in the Laurentian Great Lakes region.

METHODS

Model Description

Model Spatial Structure and Forcing Functions

The spatial structure of the model is simpler than the two-compartment vertical system of the original model application in Lake Washington (Arhonditsis and Brett 2005a, b). We considered a single compartment model representing the lake epilimnion, whereas the hypolimnion was treated as boundary conditions to emulate mass exchanges across the thermocline. The external forcing encompasses river inflows, precipitation, evaporation, solar radiation, water temperature, and nutrient loading. The reference conditions for our analysis correspond to the average epilimnetic temperature, solar radiation, vertical diffusive mixing, hydraulic and nutrient loading in Lake Washington (Arhonditsis and Brett 2005b, Brett *et al.* 2005). The hydraulic renewal rate in our hypothetical system is 0.384 $year^{-1}$. The fluvial and aerial total nitrogen inputs are $1{,}114 \times 10^3$ kg $year^{-1}$, and the exogenous total phosphorus loading contributes approximately 74.9×10^3 kg $year^{-1}$. The exogenous total organic carbon supplies in the system are $6{,}685 \times 10^3$ kg $year^{-1}$. In our analysis, the average input nutrient concentrations for the oligo-, meso-, and eutrophic environments correspond to 50 (2.9 mg TOC/L, 484 μg TN/L, and 32.5 μg TP/L), 100 (5.8 mg TOC/L, 967 μg TN/L, and 65 μg TP/L), and 200% (11.6 mg TOC/L, 1,934 μg TN/L, and 130 μg TP/L) of the reference conditions, respectively. Based on these nutrient loading scenarios, the model was run using the calibration vector presented in Arhonditsis and Brett (2005a; see their Appendix B for parameter definitions and calibration values). The simulated monthly averages provided the mean values of normal distributions with standard deviations assigned to be 15% of the monthly values for each state variable; a fraction that comprises both analytical error and interannual variability at the deeper (middle) sections of the lake. These distributions were then sampled to generate the oligo-, meso- and eutrophic datasets used for the Bayesian model calibration.

Plankton Community Structure

The ecological submodel consists of 24 state variables and simulates five elemental cycles (organic C, N, P, Si, O) as well as three phytoplankton (diatoms, green algae, and cyanobacteria) and two zooplankton (copepods and cladocerans) groups (Arhonditsis and Brett 2005a, b). The three phyto-

plankton functional groups differ with regard to their strategies for resource competition (nitrogen, phosphorus, light, temperature) and metabolic rates as well as their morphological features (settling velocity, shading effects) (Fig. 1a). Phytoplankton growth temperature dependence has an optimum level and is modeled by a function similar to a Gaussian probability curve (Cerco and Cole 1994). Phosphorus and nitrogen dynamics within the phytoplankton cells account for luxury uptake, and phytoplankton uptake rates depend on both intracellular and extracellular nutrient concentrations (Schladow and Hamilton 1997, Arhonditsis *et al*. 2002). We used Steele's equation to describe the relationship between photosynthesis and light intensity along with Beer's law to scale photosynthetically active radiation to depth (Jassby and Platt 1976). Diatoms are modeled as r-selected organisms with high maximum growth rates and higher metabolic losses, strong phosphorus and weak nitrogen competitors, lower tolerance to low light availability, low temperature optima, silica requirements, and high sinking velocities. By contrast, cyanobacteria are modeled as K-strategists with low maximum growth and metabolic rates, weak P and strong N competitors, higher tolerance to low light availability, low settling velocities, and high temperature optima. The parameterization of the third functional group (labelled as "Green Algae") aimed to provide an intermediate competitor and more realistically depict the continuum between diatom- and cyanobacteria-dominated phytoplankton communities.

The two zooplankton functional groups (cladocerans and copepods) differ with regard to their grazing rates, food preferences, selectivity strategies, elemental somatic ratios, vulnerability to predators, and temperature requirements (Arhonditsis and Brett 2005a, b). Cladocerans are modeled as filter-feeders with an equal preference among the four food-types (diatoms, green algae, cyanobacteria, detritus), high maximum grazing rates and metabolic losses, lower half saturation for growth efficiency, high temperature optima and high sensitivity to low temperatures, low nitrogen and high phosphorus content. In contrast, copepods are characterized by lower maximum grazing and metabolic rates, capability of selecting on the basis of food quality, higher feeding rates at low food abundance, slightly higher nitrogen and much lower phosphorus content, lower temperature optima with a wider temperature tolerance. Fish predation on cladocerans is modeled by a sigmoid function, while a hyperbolic form is adopted for copepods (Edwards and Yool 2000). Both forms exhibit a plateau at high zooplankton concentrations representing satiation of the fish predation, e.g., the fish can only process a certain number of food items per unit time or there is a maximum limit on predator density caused by direct interference among the predators themselves. The S-shaped curve, however, is more appropriate for reproducing the tight connection between planktivorous fish and large *Daphnia* adults at higher zooplankton densities, due to fish specialization (learning ability of fish to capture large animals) or lack of escape behavior of the prey (Lampert and Sommer 1997).

Carbon Cycle

The inorganic carbon required for algal photosynthesis is assumed to be in excess and thus is not explicitly modeled. Dissolved organic carbon (DOC) and particulate organic carbon (POC) are the two carbon state variables considered by the model (Fig. 1b). Phytoplankton basal metabolism, zooplankton basal metabolism and egestion of excess carbon during zooplankton feeding release particulate and dissolved organic carbon in the water column. A fraction of the particulate organic carbon undergoes first-order dissolution to dissolved organic carbon, while another fraction settles to the sediment. Particulate organic carbon is grazed by zooplankton (detrivory), dissolved organic carbon is lost through a first-order denitrification and respiration during heterotrophic activity.

Nitrogen Cycle

There are four nitrogen forms considered by the model: nitrate (NO_3), ammonium (NH_4), dissolved organic nitrogen (DON), particulate organic nitrogen (PON) (Fig. 1c). Both ammonium and nitrate are utilized by phytoplankton during growth and Wroblewski's model (1977) was used to describe ammonium inhibition of nitrate uptake. Phytoplankton basal metabolism, zooplankton basal metabolism, and egestion of excess nitrogen during zooplankton feeding release ammonium and organic nitrogen in the water column. A fraction of the particulate organic nitrogen hydrolyzes to dissolved organic nitrogen. Dissolved organic nitrogen is mineralized to ammonium. In an oxygenated water column, ammonium is oxidized to nitrate through nitrification and its kinetics are modeled as a function of available ammonium, dissolved oxy-

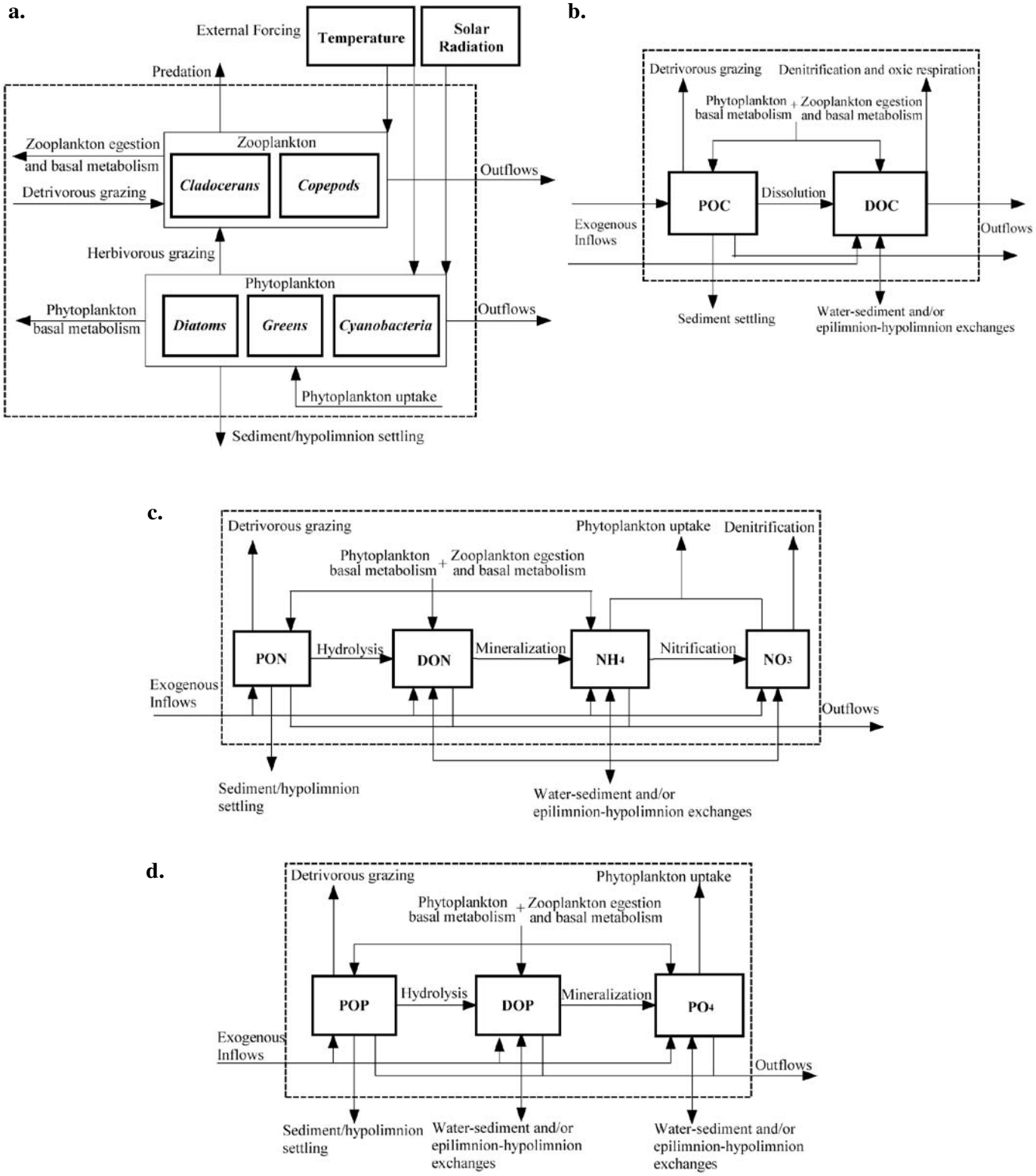

FIG. 1. The structure of the complex aquatic biogeochemical model. Arrows indicate flows of matter through the system: a. plankton submodel; b. carbon cycle; c. nitrogen cycle; d. phosphorus cycle.

gen, temperature and light (Cerco and Cole 1994, Tian *et al*. 2001). During anoxic conditions, nitrate is lost as nitrogen gas through denitrification.

Phosphorus Cycle

Three phosphorus state variables were considered in the model: phosphate (PO_4), dissolved organic phosphorus (DOP), and particulate organic phosphorus (POP) (Fig. 1d). Phytoplankton uptakes phosphate and redistributes the three forms of phosphorus through basal metabolism. Zooplankton basal metabolism and egestion of excess phosphorus during feeding release phosphate and dissolved/particulate organic phosphorus. Particulate organic phosphorus can be hydrolyzed to dissolved organic phosphorus, and another fraction settles to the sediment. Dissolved organic phosphorus is mineralized to phosphate through a first-order reaction.

Bayesian Framework

Statistical Formulation

Our presentation examines a statistical formulation founded on the assumption that 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. This formulation aims to combine field observations with simulation model outputs to update the uncertainty of model parameters, and then use the calibrated model to give predictions (along with uncertainty bounds) of the natural system dynamics. An observation i for the state variable j, y_{ij}, can be described as:

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

$$g(\theta, x_i, y_0, \delta_j) \sim N(f(\theta, x_i, y_0), \sigma_j^2)$$

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, forcing functions) describing the environmental conditions, the vector θ is a time independent set of the calibration model parameters, y_0 corresponds to the vector of the concentrations of the twenty four state-variables at the initial time point t_0 (initial conditions), the stochastic term δ_j accounts for the discrepancy between the model and the natural system, ε_{ij} denotes the observation (measurement) error that is usually assumed to be independent and identically distributed following a Gaussian distribution, and $g(\theta, x_i, y_0, \delta_j)$ represents a normally distributed variable with first and second order moments based on the model predictions and the time independent model structural error σ_j^2. In this study, as a result of the scheme followed to generate the three datasets, we assumed a multiplicative measurement error with standard deviations proportional (15%) to the average monthly values for each state variable (Van Oijen *et al*. 2005). With this assumption, the likelihood function (see Glossary of Terms) will be:

$$p\left(y \middle| f(\theta, x, y_0)\right) = \prod_{j=1}^{m} (2\pi)^{-n/2} \left|\sum_{Totj}\right|^{-1/2} \tag{2}$$

$$\exp\left[-\frac{1}{2}\left[y_j - f_j(\theta, x, y_0)\right]^T \sum_{Totj}^{-1} \left[y_j - f_j(\theta, x, y_0)\right]\right]$$

$$\sum_{Totj} = \sum_{\delta j} + \sum_{\varepsilon j} \tag{3}$$

where m and n correspond to the number of state variables (m = 24) 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 $f_j(\theta, x, y_0) = [f_{1j}(\theta, x_1, y_0),\ldots, f_{nj}(\theta, x_n, y_0)]^T$ correspond to the vectors of the field observations and model predictions for the state variable j; $\Sigma_{\delta j} = I_n \cdot \sigma_j^2$ corresponds to the stochastic term of the model; and $\Sigma_{\varepsilon j} = I_n \cdot (0.15)^2 \cdot y_j^T \cdot y_j$. In the context of the Bayesian statistical inference, the posterior density of the parameters θ and the initial conditions of the twenty four state variables y_0 given the observed data y is defined as:

$$p\left(\theta, y_0, \sigma^2 \middle| y\right) = \frac{p\left(y \middle| f(\theta, x, y_0, \sigma^2)\right) p(\theta) p(y_0) p(\sigma^2)}{\int\int\int p\left(y \middle| f(\theta, x, y_0, \sigma^2)\right) p(\theta) p(y_0) p(\sigma^2) d\theta dy_0 d\sigma^2} \propto p\left(y \middle| f(\theta, x, y_0, \sigma^2)\right) p(\theta) p(y_0) p(\sigma^2) \tag{4}$$

$p(\theta)$ is the prior density of the model parameters θ and $p(y_0)$ is the prior density of the initial conditions of the twenty 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; the prior densities $p(\sigma_j^2)$ were based on the

conjugate inverse-gamma distribution (Gelman *et al*. 1995). Thus, the resulting posterior distribution for θ, $y_{0,}$ and σ^2 is:

$$p\left(\theta, y_0, \sigma^2 | y\right) \propto \prod_{j=1}^{m} (2\pi)^{-n/2} \left|\sum {}_{Totj}\right|^{-1/2}$$
$$\exp\left[-\frac{1}{2}\left[y_j - f_j(\theta, x, y_0)\right]^T \sum {}_{Totj}^{-1}\left[y_j - f_j(\theta, x, y_0\right]\right]$$
$$\times(2\pi)^{-l/2} \left|\sum {}_{\theta}\right|^{-1/2} \prod_{k=1}^{l} \frac{1}{\theta_k} \qquad (5)$$
$$\exp\left[-\frac{1}{2}\left[\log\theta - \theta_0\right]^T \sum {}_{\theta}^{-1}\left[\log\theta - \theta_0\right]\right]$$
$$\times(2\pi)^{-m/2} \left|\sum {}_{y0}\right|^{-1/2} \exp\left[-\frac{1}{2}\left[y_0 - y_{0m}\right]^T \sum {}_{y0}^{-1}\left[y_0 - y_{0m}\right]\right]$$
$$\times\prod_{j=1}^{M} \frac{\beta_j^{\alpha j}}{\Gamma(\alpha_j)} \sigma_j^{-2(\alpha_j+1)} \exp\left(-\frac{\beta_j}{\sigma_j^2}\right)$$

where l is the number of the model parameters θ used for the model calibration ($l = 35$); θ_0 indicates the vector of the mean values of θ in logarithmic scale; $\Sigma_\theta = I_l \cdot \sigma_\theta^T \cdot \sigma_\theta$ and $\sigma_\theta = [\sigma_{\theta l}, \ldots, \sigma_{\theta l}]^T$ corresponds to the vector of the shape parameters of the l lognormal distributions (standard deviation of *log* θ); the vector $y_{0m} = [y_{1,1}, \ldots, y_{1,24}]^T$ corresponds to the January values of the twenty four state variables; $\Sigma_{y0} = I_m \cdot (0.15)^2 \cdot y_{0m}^T \cdot y_{0m}$; α_j (= 0.01) and β_j (= 0.01) correspond to the shape and scale parameters of the m non-informative inverse-gamma distributions used in this analysis.

Prior Parameter Distributions

The calibration vector consists of the 35 most influential parameters as identified from an earlier sensitivity analysis of the model (Arhonditsis and Brett 2005a). The prior parameter distributions reflect the existing knowledge (field observations, laboratory studies, literature information, and expert judgment) on the relative plausibility of their values. For example, based on the previous characterization of the three functional groups, we assigned probability distributions that represent their differences in growth and storage strategies, basal metabolism, nitrogen and phosphorus kinetics, light and temperature requirements, and settling velocity. In this study, we used the following protocol to formulate the parameter distributions: i) we identified the global (not the group-specific) minimum and maximum values for each parameter from the pertinent literature; ii) we partitioned the original parameter space into three subregions reflecting the functional properties of the phytoplankton groups; and then iii) we assigned lognormal distributions parameterized such that 98% of their values were lying within the identified ranges (Steinberg *et al*. 1997). The group-specific parameter spaces were also based on the calibration vector presented during the model application in Lake Washington (Arhonditsis and Brett 2005a). For example, the identified range for the maximum phytoplankton growth rate was 1.0–2.4 day^{-1}, while the three subspaces were 2.2 ± 0.2 day^{-1} for diatoms (calibration value ± literature range), 1.8 ± 0.2 day^{-1} for greens and 1.3 ± 0.3 day^{-1} for cyanobacteria. We then assigned lognormal distributions formulated such that 98% of their values were lying within the specified ranges, i.e., $growth_{max(diat)} \sim \Lambda(2.19, 1.040)$, $growth_{max(greens)} \sim \Lambda(1.79, 1.049)$, $growth_{max(cyan)} \sim \Lambda(1.26, 1.106)$. The prior distributions of all the parameters of the model calibration vector are presented in Table 1.

Numerical Approximations for Posterior Distributions

Sequence of realizations from the posterior distribution of the model were obtained using Markov chain Monte Carlo (MCMC) simulations (Gilks *et al*. 1998). We used the general normal-proposal Metropolis algorithm coupled with an ordered overrelaxation to control the serial correlation of the MCMC samples (Neal 1998). In this study, we present results using two parallel chains with starting points: (i) a vector that consists of the mean values of the prior parameter distributions, and (ii) the calibration vector of the application Lake Washington. 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 Spiegelhalter *et al*. 2003) for all the parameters was less than 5% of the sample standard deviation. Our framework was implemented in the WinBUGS Differential Interface (WBDiff); an interface that allows numerical solution of systems of ordinary differential equations within the WinBUGS software.

TABLE 1. Prior and posterior parameter distributions in three trophic states: Λ– lognormal distribution, $\theta \sim \Lambda(\mu^*, \sigma^*)$ is a mathematical expression meaning that θ is lognormally distributed, μ^* and σ^* correspond to the median and multiplicative standard deviation.

Parameters	Prior	Oligotrophic	Mesotrophic	Eutrophic
$bm_{ref(clad)}$	Λ(0.0495, 1.161)	Λ(0.0491, 1.236)	Λ(0.0490, 1.239)	Λ(0.0491, 1.241)
$bm_{ref(cop)}$	Λ(0.0442, 1.181)	Λ(0.0441, 1.271)	Λ(0.0438, 1.271)	Λ(0.0444, 1.265)
$bm_{ref(cyan)}$	Λ(0.0775, 1.116)	Λ(0.0774, 1.168)	Λ(0.0789, 1.163)	Λ(0.0808, 1.162)
$bm_{ref(diat)}$	Λ(0.0980, 1.091)	Λ(0.0978, 1.144)	Λ(0.0951, 1.125)	Λ(0.0946, 1.120)
$bm_{ref(green)}$	Λ(0.0775, 1.116)	Λ(0.0760, 1.170)	Λ(0.0753, 1.164)	Λ(0.0753, 1.163)
$ef_{2(clad)}$	Λ(18.3, 1.123)	Λ(18.3, 1.183)	Λ(18.3, 1.181)	Λ(18.1, 1.183)
$ef_{2(cop)}$	Λ(19.4, 1.116)	Λ(19.3, 1.174)	Λ(19.3, 1.172)	Λ(19.4, 1.166)
$growth_{max(cyan)}$	Λ(1.26, 1.106)	Λ(1.29, 1.155)	Λ(1.28, 1.158)	Λ(1.22, 1.145)
$growth_{max(diat)}$	Λ(2.19, 1.040)	Λ(2.23, 1.050)	Λ(2.24, 1.049)	Λ(2.22, 1.055)
$growth_{max(greens)}$	Λ(1.79, 1.049)	Λ(1.80, 1.070)	Λ(1.80, 1.073)	Λ(1.81, 1.070)
$grazing_{max(clad)}$	Λ(0.837, 1.080)	Λ(0.837, 1.118)	Λ(0.839, 1.115)	Λ(0.844, 1.121)
$grazing_{max(cop)}$	Λ(0.490, 1.091)	Λ(0.489, 1.134)	Λ(0.477, 1.125)	Λ(0.490, 1.139)
$KC_{refdissolution}$	Λ(0.00200, 2.691)	Λ(0.00194, 2.573)	Λ(0.00198, 2.588)	Λ(0.00206, 2.643)
$K_{eddyref}$	Λ(0.0316, 1.218)	Λ(0.0351, 1.277)	Λ(0.0325, 1.340)	Λ(0.0322, 1.277)
$K_{EXTback}$	Λ(0.265, 1.084)	Λ(0.256, 1.106)	Λ(0.244, 1.075)	Λ(0.252, 1.097)
$K_{EXTchla}$	Λ(0.0200, 1.347)	Λ(0.0187, 1.489)	Λ(0.0169, 1.424)	Λ(0.0173, 1.452)
$KN_{(cyan)}$	Λ(22.9, 1.200)	Λ(22.8, 1.308)	Λ(22.9, 1.298)	Λ(23.0, 1.306)
$KN_{(diat)}$	Λ(64.2, 1.069)	Λ(64.1, 1.101)	Λ(64.1, 1.101)	Λ(64.2, 1.101)
$KN_{(greens)}$	Λ(43.9, 1.102)	Λ(43.9, 1.151)	Λ(43.9, 1.150)	Λ(43.7, 1.149)
$KN_{refdissolution}$	Λ(0.00200, 2.691)	Λ(0.00201, 2.663)	Λ(0.00199, 2.613)	Λ(0.00195, 2.594)
$KN_{refmineral}$	Λ(0.00775, 1.503)	Λ(0.00884, 1.622)	Λ(0.00594, 1.559)	Λ(0.00691, 1.716)
$KP_{(cyan)}$	Λ(19.4, 1.116)	Λ(19.2, 1.174)	Λ(19.7, 1.168)	Λ(19.5, 1.174)
$KP_{(diat)}$	Λ(5.66, 1.161)	Λ(5.28, 1.216)	Λ(5.36, 1.226)	Λ(5.46, 1.235)
$KP_{(greens)}$	Λ(10.6, 1.128)	Λ(10.4, 1.187)	Λ(10.3, 1.187)	Λ(10.4, 1.188)
$KP_{refdissolution}$	Λ(0.00200, 2.691)	Λ(0.00202, 2.604)	Λ(0.00198, 2.603)	Λ(0.00202, 2.668)
$KP_{refmineral}$	Λ(0.0245, 1.470)	Λ(0.0220, 1.644)	Λ(0.0235, 1.691)	Λ(0.0235, 1.716)
$KSi_{(diat)}$	Λ(40.0, 1.347)	Λ(39.7, 1.542)	Λ(39.8, 1.536)	Λ(39.8, 1.527)
$KSi_{refdissolution}$	Λ(0.00200, 2.691)	Λ(0.00198, 2.631)	Λ(0.00197, 2.613)	Λ(0.00194, 2.533)
$KZ_{(clad)}$	Λ(114, 1.058)	Λ(114, 1.087)	Λ(114, 1.087)	Λ(113, 1.085)
$KZ_{(cop)}$	Λ(93.8, 1.071)	Λ(93.6, 1.104)	Λ(94.5, 1.104)	Λ(93.3, 1.100)
$pred_1$	Λ(0.141, 1.161)	Λ(0.139, 1.238)	Λ(0.138, 1.233)	Λ(0.136, 1.224)
$pred_2$	Λ(34.6, 1.266)	Λ(36.1, 1.400)	Λ(35.5, 1.412)	Λ(39.4, 1.330)
$V_{settling(cyan)}$	Λ(0.0224, 1.413)	Λ(0.0205, 1.590)	Λ(0.0224, 1.605)	Λ(0.0232, 1.610)
$V_{settling(diat)}$	Λ(0.316, 1.106)	Λ(0.289, 1.112)	Λ(0.275, 1.072)	Λ(0.293, 1.118)
$V_{settling(greens)}$	Λ(0.245, 1.091)	Λ(0.237, 1.128)	Λ(0.231, 1.108)	Λ(0.235, 1.120)

RESULTS

The MCMC sequences of the three applications of the model converged rapidly (≈ 5,000 iterations) and the statistics reported were based on the last 25,000 draws by keeping every 4th iteration (thin = 4). The uncertainty underlying the values of the 35 model parameters after the MCMC sampling is depicted on the respective marginal posterior distributions (Table 1 and Fig. 2). Generally, the moments of the posterior parameter distributions indicate that the knowledge gained for the 35 parameters after the Bayesian updating of the complex eutrophication model was fairly limited. [It should be noted that for the sake of consistency all the parameter posteriors were presented as lognormal distributions, although in several cases the shape is better approximated by a uniform distribution.] Namely, most of the calibration parameters were characterized by minor or no shifts of their central tendency relative to the prior assigned values, such as the half saturation constants for nitrogen uptake ($KN_{(i)}$; *i= diatoms, greens, cyanobacteria*), the half saturation constants for grazing ($KZ_{(j)}$; *j= cladocerans, copepods*), and the half saturation constants for growth efficiency ($ef_{2(j)}$; *j= cladocerans, copepods*). Nonetheless, there were parameters with moderate shifts of their posterior mean values; characteristic examples were the ni-

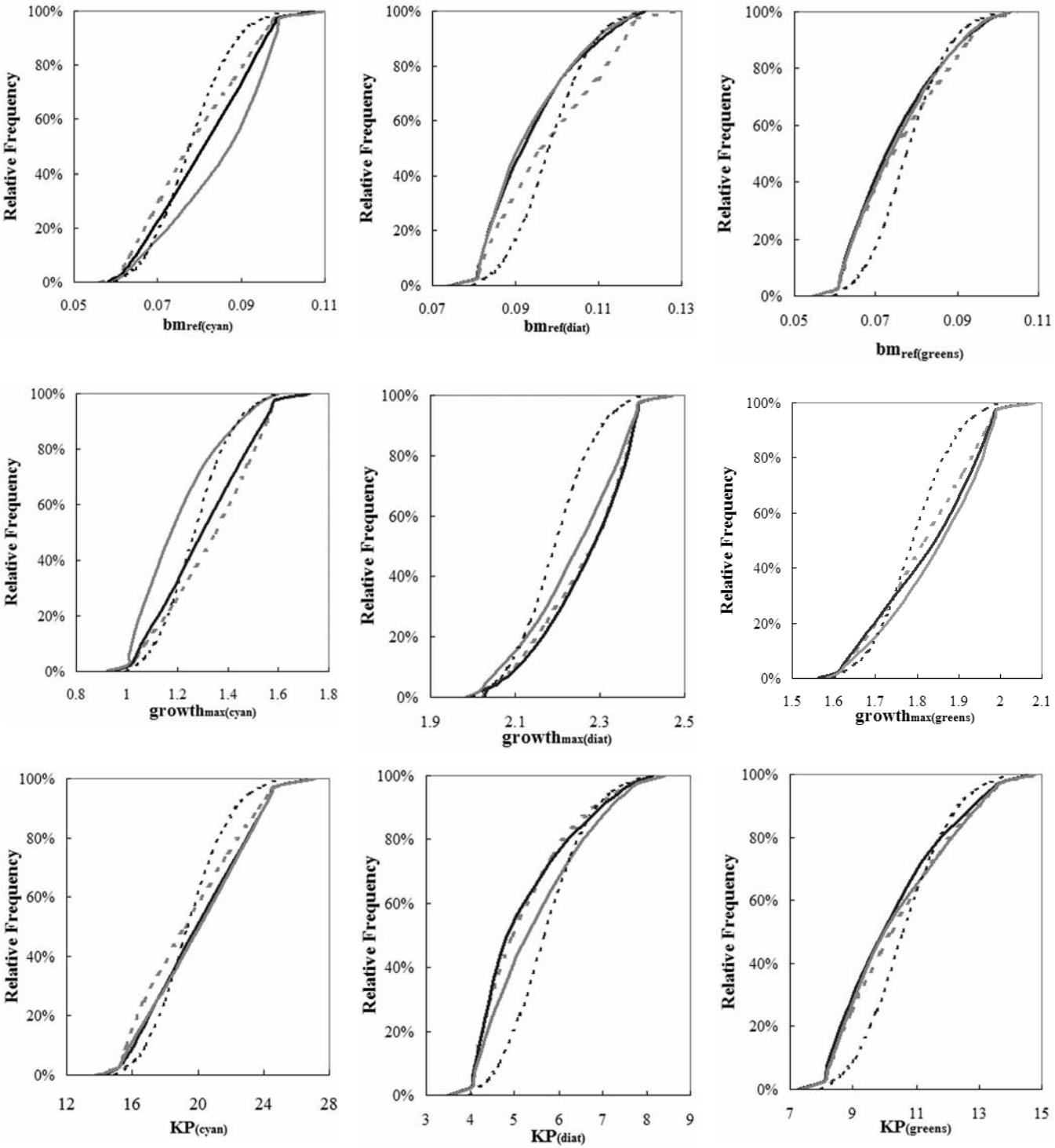

FIG. 2. Prior (thin black dashed lines) and posterior (eutrophic environment: thick grey lines, mesotrophic environment: thick black lines, and oligotrophic environment: thick grey dashed lines,) cumulative distributions of the aquatic biogeochemical model.

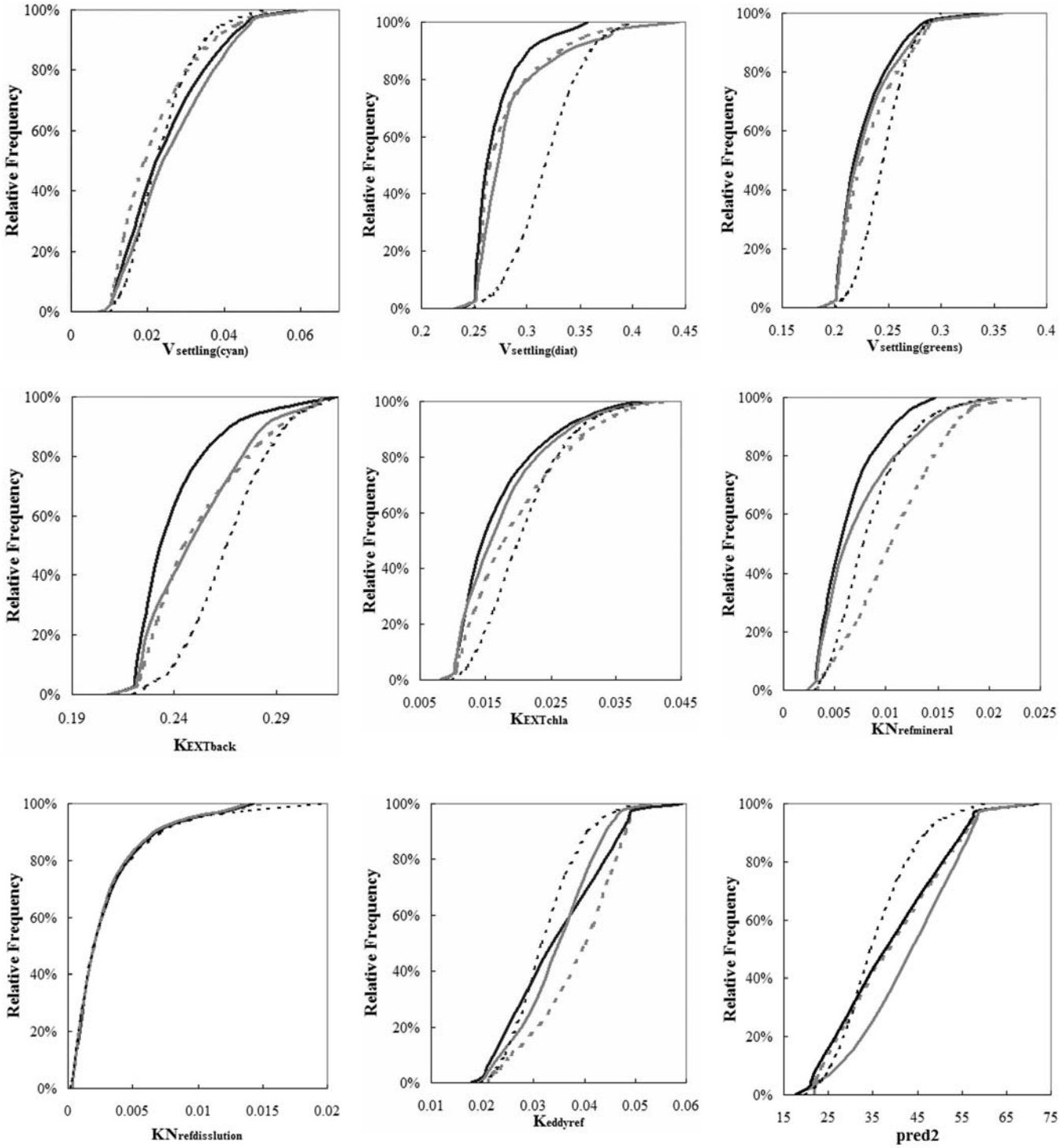

FIG. 2. (Continued).

trogen mineralization rate ($KN_{refmineral}$) with relative percentage changes of 14, 23, and 11% in the oligo-, meso-, and eutrophic environments, respectively; the light attenuation coefficient for chlorophyll ($K_{EXTchla}$) with 6, 15, and 14% relative changes in the three nutrient enrichment conditions; settling velocity for diatoms ($V_{settling(diat)}$) with 9, 13, and 7% relative shifts. Furthermore, the vast majority of the posterior standard deviations increased or remained unaltered relative to the prior assigned values, and several parameter posteriors were almost uniformly distributed within the specified ranges prior to the model calibration. Notable exceptions were the dissolution/hydrolysis rates for particulate carbon ($KC_{refdissolution}$), nitrogen ($KN_{refdissolution}$), phosphorus ($KP_{refdissolution}$), and silica ($KSi_{refdissolution}$) with approximately 2–6% relative decrease of the respective standard deviations. The standard deviation of the diatom settling velocity ($V_{settling(diat)}$) was also reduced by 3% in the mesotrophic state.

The comparison between the observed and posterior predictive monthly distributions for the three trophic states indicates that the eutrophication model combined with the Bayesian calibration scheme provides an accurate representation of the system dynamics. In the oligotrophic environment, the observed monthly values were included within the 95% credible intervals of the model predictions throughout the simulation period, while the median values of model predictions closely matched the observed patterns (Fig. 3). In a similar manner, all the observed values of the dataset representing the mesotrophic conditions were included within the 95% credible intervals, although the median model predictions slightly underestimated the spring biomass peaks of three phytoplankton groups (Fig. 4). In the eutrophic scenario, the model closely reproduced the summer prey-predator oscillations between cladocerans and the three phytoplankton groups and also accurately simulated the nutrient dynamics, i.e., total nitrogen, nitrate, ammonium, total phosphorus, and phosphate (Fig. 5). However, the central tendency and uncertainty bounds of the copepod biomass predictive distribution failed to capture the late-spring peak, while the upper (97.5%) and lower (2.5%) uncertainty boundaries showed convexo-convex shape during the same period.

The model performance for each trophic state was evaluated by three measures of fit: root mean squared error (RMSE), relative error (RE) and average error (AE) (Table 2). These comparisons aimed to assess the goodness-of-fit between the medians of the predictive distributions and the observed values. The application of the model to the oligotrophic environment was characterized by the lowest RE values (1.19–10.6%), while the mesotrophic and eutrophic scenarios resulted in moderate (3.37–13.6%), and relatively larger RE values (6.03–21.2%), respectively. We also highlight the fairly high RE values for cyanobacteria and copepod biomass in the eutrophic environment, whereas total nitrogen and dissolved oxygen had consistently low REs in the three nutrient loading scenarios. The average error is a measure of aggregate model bias, though values near zero can be misleading because negative and positive discrepancies can cancel each other. In most cases, we found that the medians of the state variable predictive distributions underestimated the observed levels, whereas dissolved oxygen was overestimated with an AE value of 0.482, 0.356, and 0.628 mg L^{-1} in the oligo-, meso-, and eutrophic environment, respectively. The root mean square error is another measure of the model prediction accuracy that overcomes the shortcoming of the average error by considering the magnitude rather than the direction of each difference. The RMSE for the copepod biomass increased across the trophic gradient examined from 5.19 μg C L^{-1} in the oligotrophic to 13.2 and 48.3 μg C L^{-1} in the meso- and eutrophic datasets, respectively. We also note the approximately 0.5 μg chla L^{-1} mean discrepancy between the predictive medians and the observed cyanobacteria biomass values.

The seasonally invariant error terms (σ_j) delineate a constant zone around the model predictions for the 24 state variables that accounts for the discrepancy between the model simulation and the natural system dynamics (Table 3). The majority of the discrepancy terms increased as we move from the oligotrophic to the eutrophic state, providing evidence that these terms play an important role in accommodating the increased intra-annual variability of the meso- and eutrophic datasets. On the other hand, the error terms associated with the phytoplankton intracellular nutrient storage (e.g., $\sigma_{N,\ P(i);\ i=\ diatoms,\ greens,\ cyanobacteria}$, and $\sigma_{Si(diatoms)}$) were characterized by similar mean and standard deviation values across the trophic gradient examined. Finally, high coefficients of variation (standard deviation/mean) were found for the dissolved oxygen, dissolved organic carbon, and dissolved silica error terms.

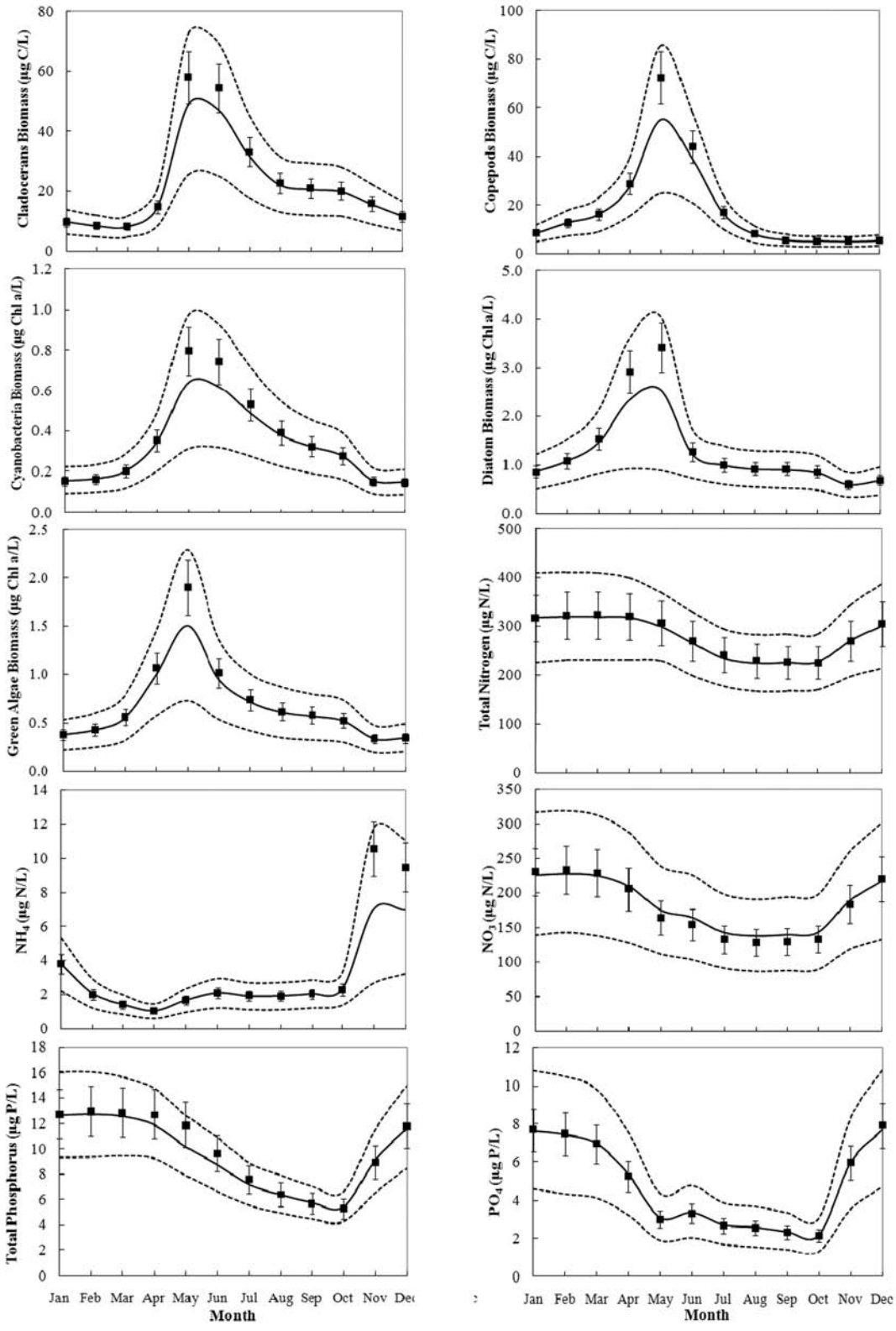

FIG. 3. Comparison between the observed and posterior predictive monthly distributions for 10 water quality variables based on Markov chain Monte Carlo posterior samples from the model application in the oligotrophic environment. Solid line corresponds to the median value of model prediction and dashed lines correspond to the 2.5 and 97.5% uncertainty bounds. The square dots represent the observed data, while the error bars correspond to the measurement error.

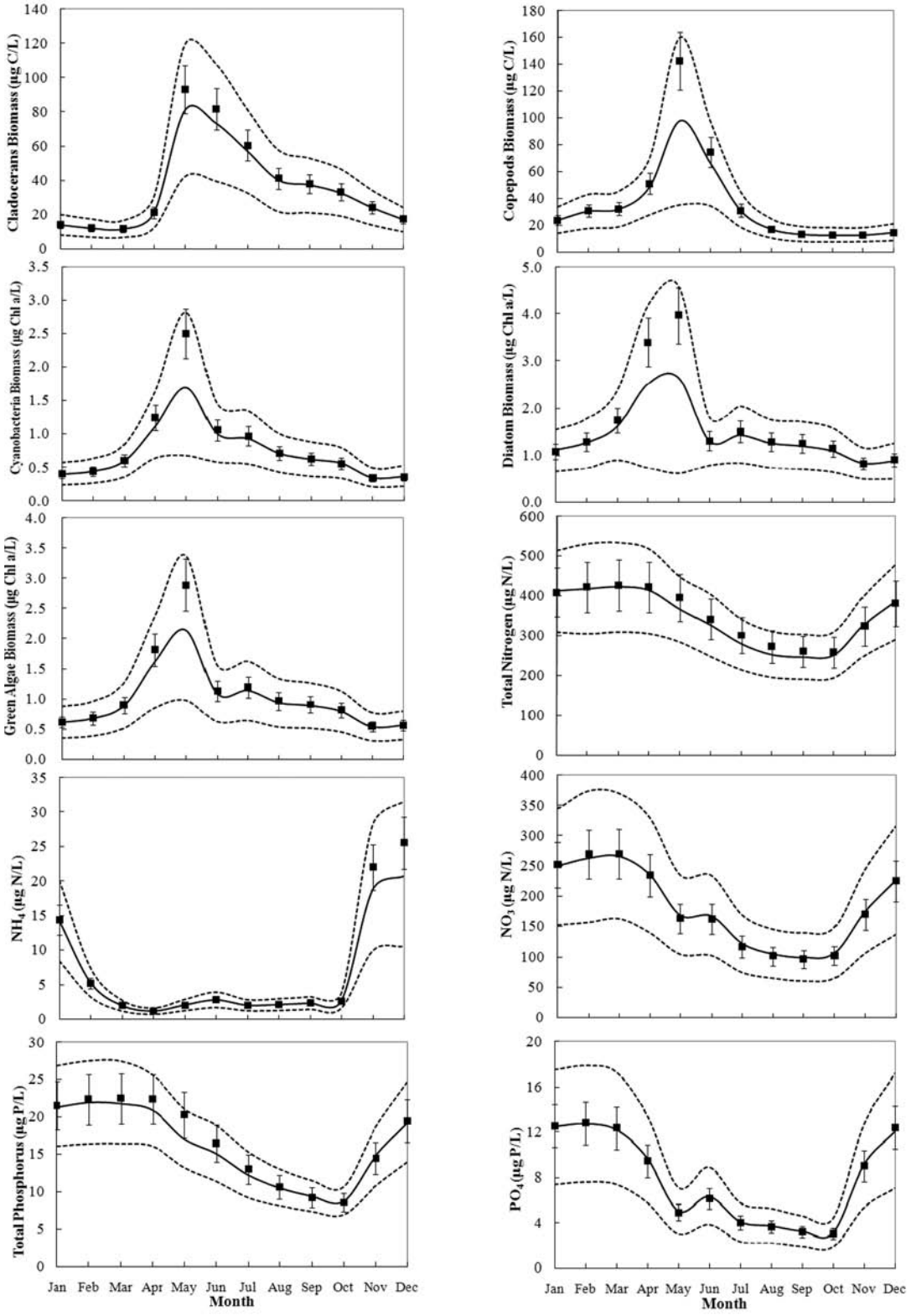

FIG. 4. Comparison between the observed and posterior predictive monthly distributions for 10 water quality variables based on Markov chain Monte Carlo posterior samples from the model application in the mesotrophic environment. Solid line corresponds to the median value of model prediction and dashed lines correspond to the 2.5 and 97.5% uncertainty bounds. The square dots represent the observed data, while the error bars correspond to the measurement error.

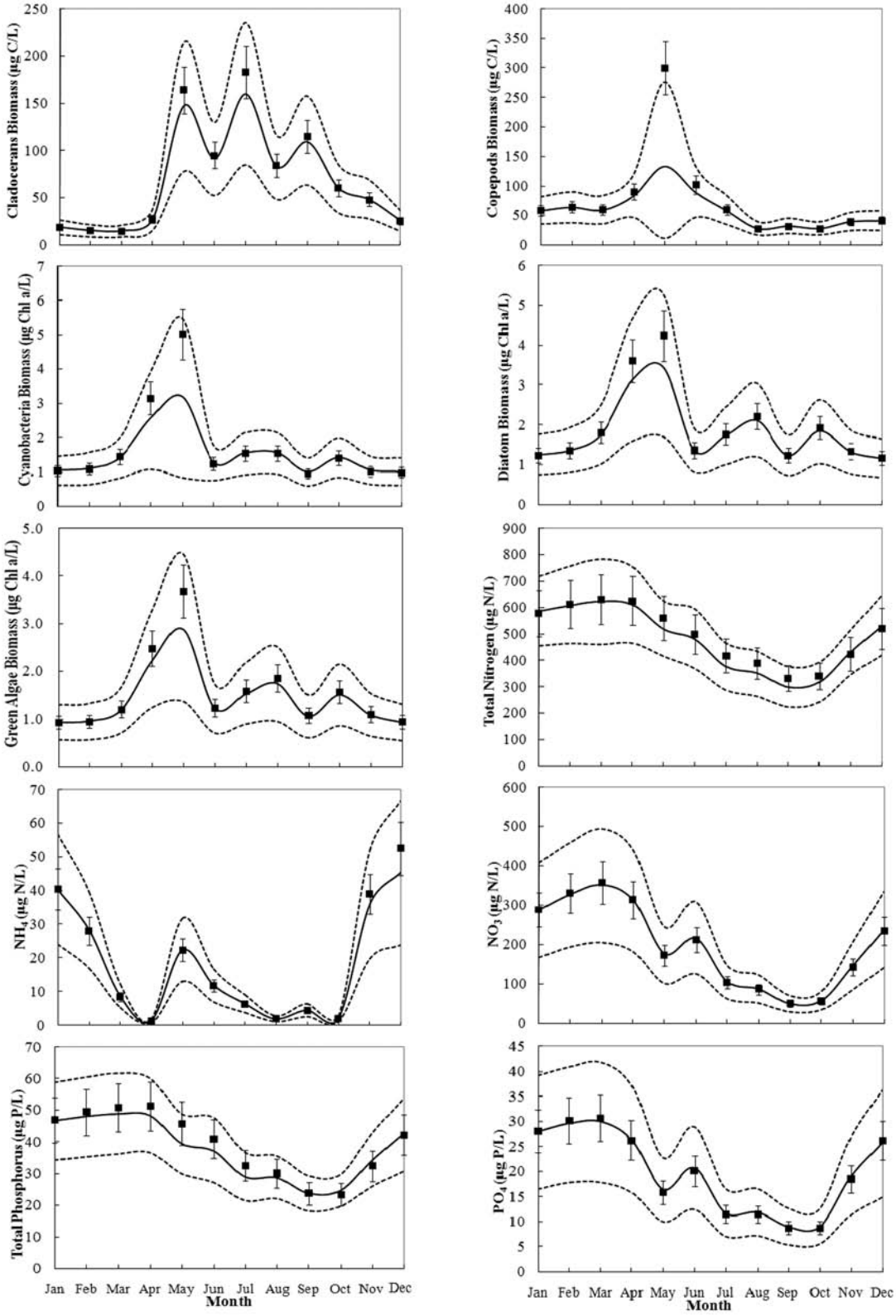

FIG. 5. Comparison between the observed and posterior predictive monthly distributions for 10 water quality variables based on Markov chain Monte Carlo posterior samples from the model application in the eutrophic environment. Solid line corresponds to the median value of model prediction and dashed lines correspond to the 2.5 and 97.5% uncertainty bounds. The square dots represent the observed data, while the error bars correspond to the measurement errors.

TABLE 2. Goodness-of-fit statistics for the model state variables in three trophic states*.

	Oligotrophic			Mesotrophic			Eutrophic		
State Variables	RMSE	RE	AE	RMSE	RE	AE	RMSE	RE	AE
Green Algae Biomass (μg Chl a/L)	0.118	7.03%	–0.050	0.223	8.49%	–0.092	0.251	7.63%	–0.117
Diatom Biomass (μg Chl a/L)	0.307	10.4%	–0.139	0.467	13.6%	–0.215	0.275	7.17%	–0.139
Cyanobacteria Biomass (μg Chl a/L)	0.059	8.26%	–0.028	0.235	10.7%	–0.082	0.552	12.8%	–0.188
Copepod Biomass (μg C/L)	5.19	10.6%	–2.00	13.2	12.6%	–4.74	48.3	21.2%	–15.6
Cladoceran Biomass (μg C/L)	3.41	7.04%	–1.62	4.40	5.92%	–2.20	8.42	6.03%	–4.23
Total Silica (mg Si/L)	0.097	7.50%	0.019	0.136	8.25%	–0.0085	0.222	8.61%	–0.0030
Total Nitrogen (μg N/L)	4.06	1.19%	–2.77	14.4	3.37%	–9.48	45.4	7.64%	–12.5
Total Phosphorus (μg P/L)	0.627	4.16%	–0.350	1.17	4.62%	–0.648	4.74	9.81%	–1.16
Dissolved Oxygen (mg DO/L)	0.655	4.92%	0.482	0.629	5.04%	0.356	0.763	6.37%	0.628

* RMSE – Root Mean Square Error
RE – Relative Error
AE – Average Error

Exceedance Frequency and Confidence of Compliance with Water Quality Standards

The MCMC posterior samples were also used to examine the exceedance frequency and confidence of compliance with different water quality standards under the three nutrient loading scenarios. For illustration purposes, we selected three water quality variables of management interest, i.e., chlorophyll *a* concentration, total phosphorus, and percentage cyanobacteria contribution to the total phytoplankton biomass, and then specified their threshold values (numerical criteria) at 5 μg Chl *a* L^{-1}, 25 μg TP L^{-1}, and 30%, respectively. For each iteration, we calculated the monthly predicted values and the corresponding probabilities of exceeding the three water quality criteria. The latter probabilities were calculated as follows:

$$p = P(c > c' \mid \theta, x, y_0) = 1 - F\left(\frac{c' - g(\theta, x, y_0, \delta)}{\sigma_\varepsilon}\right) \qquad (6)$$

where p is the probability of the response variable exceeding a numerical criterion c', given values of θ, x, and y_0, σ_ε is the measurement error/within-month variability, and $F(\cdot)$ is the value of the cumulative standard normal distribution. The monthly predicted values along with the calculated exceedance frequencies were then averaged over the summer stratified period (June-September). The distribution of these statistics across the posterior space (12,500 MCMC samples) can be used to assess the expected exceedance frequency and the confidence of compliance with the three water quality standards, while accounting for the uncertainty that stems from the model parameter uncertainty. It should be noted that the exceedance frequency is not necessarily normally distributed, especially since this value is calculated as the average over the stratified period (Borsuk *et al.* 2002).

In our example, no violations of the 5 μg Chl *a* L^{-1} numerical criterion are predicted in the oligo- and mesotrophic scenarios (Fig. 6). On the other hand, the chlorophyll *a* standard is likely to be violated in the eutrophic environment, and the corresponding expected exceedance (the mean of the distributions in Fig. 7) and confidence of compli-

TABLE 3. Markov Chain Monte Carlo posterior estimates of the mean values and standard deviations of the model discrepancies in three trophic states.

Discrepancy terms	Oligotrophic		Mesotrophic		Eutrophic	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
$\sigma_{green\ algae}$	25.8	7.32	33.9	10.3	46.0	14.9
σ_{diatoms}	38.4	15.8	35.6	21.4	58.4	19.3
$\sigma_{cyanobacteria}$	10.5	2.93	23.3	7.62	39.2	17.6
$\sigma_{N(greens)}$	0.0494	0.0111	0.0496	0.0113	0.0495	0.0114
$\sigma_{P(greens)}$	0.0434	0.0098	0.0435	0.0098	0.0436	0.0098
$\sigma_{N(diatoms)}$	0.0492	0.0113	0.0496	0.0117	0.0493	0.0116
$\sigma_{P(diatoms)}$	0.0436	0.0099	0.0438	0.0102	0.0438	0.0097
$\sigma_{Si(diatoms)}$	0.0618	0.0163	0.0612	0.0159	0.0608	0.0156
$\sigma_{N(cyanobacteria)}$	0.0498	0.0117	0.0497	0.0117	0.0496	0.0116
$\sigma_{P(cyanobacteria)}$	0.0436	0.0099	0.0436	0.0098	0.0438	0.0098
$\sigma_{copepods}$	19.3	5.25	29.7	9.53	33.2	17.0
$\sigma_{cladocerans}$	20.5	5.43	34.8	8.84	73.8	17.9
σ_{NO3}	53.4	15.9	92.7	23.5	157	35.7
σ_{NH4}	1.77	0.784	7.76	2.09	18.6	4.71
σ_{DON}	1.56	1.93	2.32	3.23	3.56	5.24
σ_{PON}	10.8	2.55	16.6	3.89	19.0	4.62
σ_{PO4}	3.00	0.726	5.09	1.25	10.2	2.58
σ_{DOP}	0.608	0.219	1.09	0.401	1.30	0.925
σ_{POP}	0.820	0.192	1.54	0.374	1.90	0.462
σ_{DOC}	10.3	19.0	26.4	44.5	48.2	112
σ_{POC}	54.3	13.1	90.7	20.8	109	25.8
σ_{DSi}	12.3	23.5	18.1	33.0	29.7	54.7
σ_{PSi}	119	30.1	232	56.4	461	116
σ_{DO}	67.0	158	87.4	177	93.8	190

ance (the proportion of the exceedance frequency distribution that lies below the EPA's 10% guideline; CC) were approximately 30 and 3.5%, respectively. This probabilistic assessment of the water quality conditions should make model results more useful to decision makers and stakeholders, because the deterministic statements are avoided and the optimal management schemes (e.g., reduction of nutrient loading) are determined by explicitly acknowledging an inevitable risk of non-attainment. Similar insights can be gained by the other two water quality criteria (total phosphorus and cyanobacteria percentage). In the eutrophic conditions, the exceedance frequency distribution of the 25 μg TP L^{-1} criterion was lying within the 30–100% range, and therefore it is nearly impossible to comply with the 10% EPA guideline. The latter conclusion can also be drawn with regards to the 30% cyanobacteria biomass criterion, although in this case a fairly low confidence of compliance also characterizes the mesotrophic state. Analogous statements can be made with other model endpoints of management interest, such as the spatiotemporal dissolved oxygen levels in systems experiencing problems of prolonged hypoxia (e.g., Lake Erie).

DISCUSSION

The water quality management usually relies on mathematical models with strong mechanistic basis, as this improves the confidence in predictions made for a variety of conditions. From an operational standpoint, the interpretation of model results should explicitly consider two sources of model error, i.e., the observed variability that is not explained by the model and the uncertainty arising

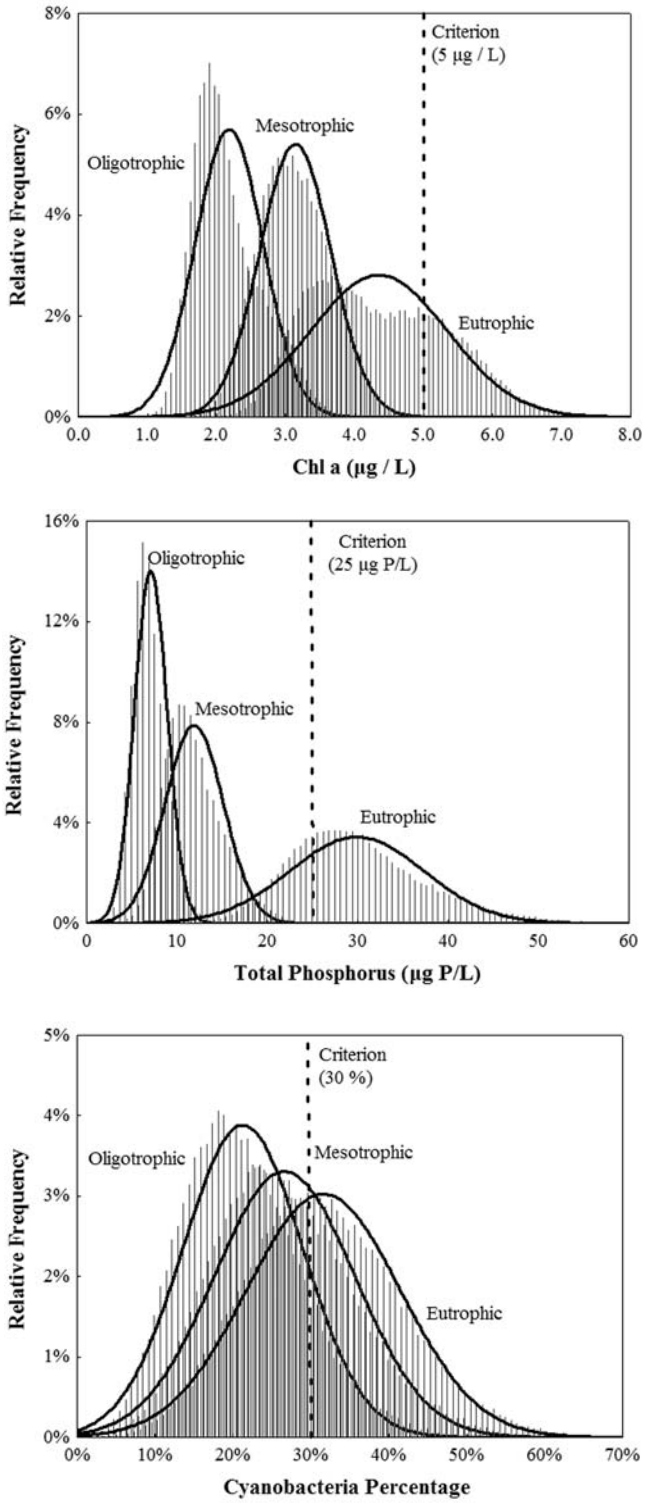

FIG. 6. Predictive distributions for water quality variables of management interest (chlorophyll a, total phosphorus, and cyanobacteria percentage) during the summer stratified period (June to September). The dashed lines correspond to the numerical criteria used to determine the frequency of violations under different trophic conditions.

from the model parameters and/or the misspecification of the model structure (Arhonditsis *et al*. 2007, Stow *et al*. 2007). In this study, we illustrated a methodological framework that can accommodate rigorous and complete error analysis, thereby allowing for the direct assessment of the frequency of water quality standard violations along with the determination of an appropriate margin of safety (Borsuk *et al*. 2002). The latter term refers to the *probability distribution of the predicted exceedance probabilities* and represents the degree of confidence that the true value of the violation frequency is below a specified value (Wild *et al*. 1996, McBride and Ellis 2001). The presentation of the model outputs as probabilistic assessment of water quality conveys significantly more information than the point predictions and is conceptually similar to the percentile-based standards proposed by the EPA-guidelines (Office of Water 1997). In this regard, our analysis also builds upon the recommendations of an earlier modeling work by Lam *et al.* (1987b), which advocated the use of probability indicators in water quality assessment in the Great Lakes area, recognizing the importance of the variability pertaining to nutrient loading and weather conditions. This type of probabilistic information is certainly more appealing to decision makers and stakeholders, as it acknowledges the knowledge gaps, the inherent uncertainty, and the interannual variability typically characterizing freshwater ecosystems (Ludwig 1996). The latter feature is particularly important in the most degraded and highly variable nearshore zones or enclosed bays/harbors in the Great Lakes. These areas are transitional zones in that they receive highly polluted inland waters from watersheds with significant agricultural, urban and/or industrial activities while mixing with offshore waters having different biological and chemical characteristics. Generally, we believe that the Bayesian calibration presented herein can be particularly useful in the context of the Great Lakes modeling, although our analysis highlighted several technical features that need to be acknowledged so as to put this framework into perspective.

As demonstrated in several recent studies (Arhonditsis *et al*. 2007, 2008a, 2008b), the inclusion of the monthly invariant stochastic terms that account for model structure imperfection resulted in a close reproduction of the epilimnetic patterns. Even though the median model predictions tend to slightly underestimate the spring plankton bloom, all the observed monthly values of the datasets rep-

resenting the three trophic states were included within the 95% credible intervals. It is important to note, however, that the updating of the model mainly changed the discrepancy error terms instead of the model input parameters; namely, the terms that reflect the model inadequacy and not the mathematical model itself were used to accommodate the temporal variability across the trophic gradient examined. The latter result does not fully satisfy the basic premise of our framework to attain realistic forecasts while gaining insight into the ecological structure (e.g., cause-effect relationships, feedback loops) underlying system dynamics. Similar results were also reported in an earlier exercise of sequential model updating (Arhonditsis *et al*. 2008a), but here the increased complexity of the model has further reduced the updating of the posterior parameter distributions. A more parsimonious statistical configuration of the model will assume a "perfect" model structure, i.e., the difference between model and lake dynamics is only caused by the observation/measurement error (Higdon *et al*. 2004, Arhonditsis *et al*. 2007). Applications of this statistical formulation resulted in narrow-shaped posterior parameter distributions but also in substantial misrepresentation of the calibration dataset (Arhonditsis *et al*. 2008a, b). Both features were attributed to the overconditioning of the parameter estimates because the lack of potential for model error tends to overestimate the information content of the observations (Beven 2006). These contradictory results highlight the pivotal role of the assumptions pertaining to model error structure and invite further examination of statistical formulations that objectively weigh the relative importance of the discrepancy terms vis-à-vis the model parameters on the calibration results. For example, future research should evaluate formulations that explicitly consider the dependence patterns of the error terms in time/space along with the covariance between measurement error and model structural error (Beven 2006, Arhonditsis *et al*. 2008b).

The determination of the model structure (and associated parameter values) that realistically represents the natural system dynamics is the basic foundation for developing robust prognostic tools (Reichert and Omlin 1997). However, most of the calibration schemes in the modeling literature have not adequately addressed the problem of uncertainty, and sometimes generate more questions than answers. Model calibration is mainly presented as an inverse solution exercise (i.e., the data for the model endpoints are used to learn something about

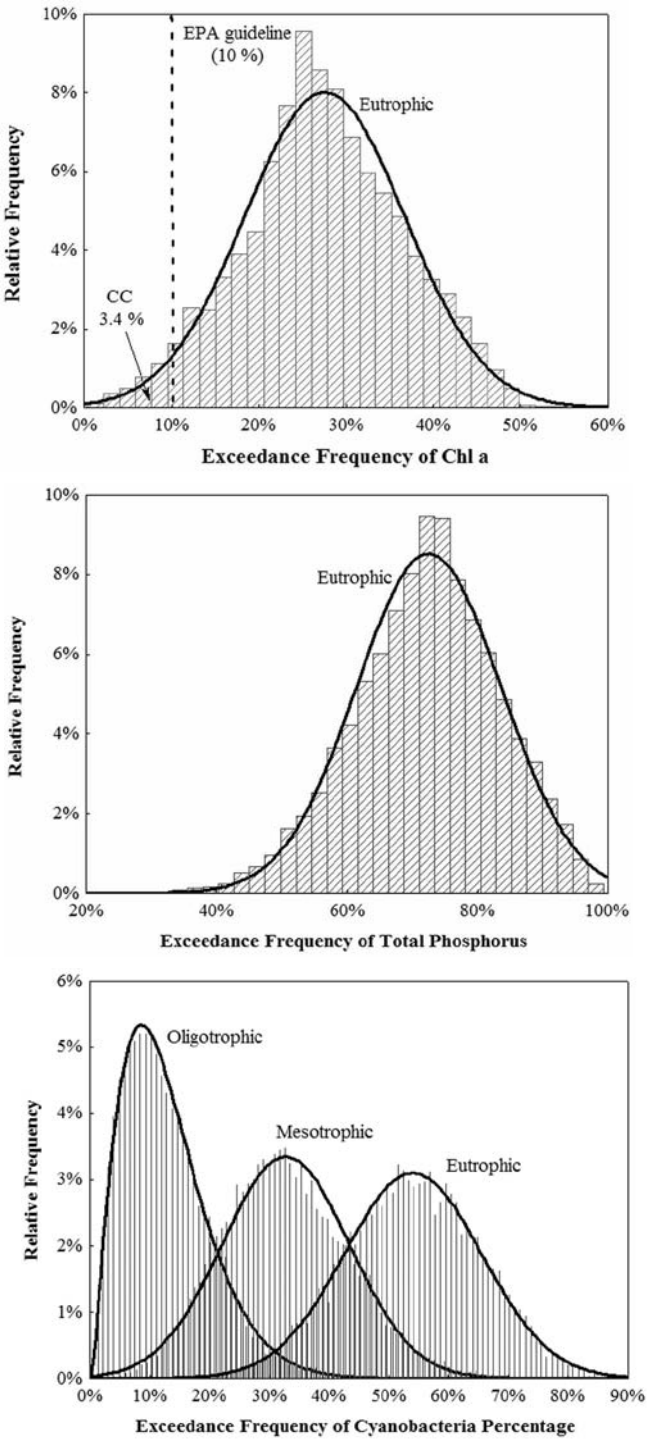

FIG. 7. The exceedance frequency of the different water quality standards (chlorophyll a: 5 µg/L, total phosphorus: 25 µg/L, and cyanobacteria percentage: 30%) during the summer stratified period (June to September) under the different trophic conditions. In these distributions, the area below the 10% cutoff point is termed the confidence of compliance (CC), and represents the probability that the true exceedance frequency is below the 10% EPA guideline.

the parameters) or as an exercise for delineating uncertainty zones around the mean predictions (Beven 1993, Beven 2001). In ecological modeling, the model parameters correspond to ecological processes for which we usually have substantial amount of information on the relative plausibility of their values (e.g., Jorgensen *et al*. 1991). Thus, it is a significant omission to ignore this knowledge and solely let the data to offer insights into the parameter marginal distributions. In this study, prior information of the magnitudes of ecological processes (based on field observations from the lake, laboratory studies, literature information, and expert judgment) was used to formulate probability distributions that reflect the relative likelihood of different values of the respective model parameters. Earlier studies have indicated that the inclusion of these informative distributions into the "prior-likelihood-posterior" update cycles of intermediate complexity models favours solutions that more realistically depict the internal structure of the system and avoid getting "good results for the wrong reasons"; the latter finding has been reported even when the mathematical models were coupled with statistical formulations that explicitly consider discrepancy error terms (Arhonditsis *et al*. 2007, 2008a, 2008b). In this analysis, however, the relatively uninformative patterns of the posterior parameter space suggest that the efficiency of this scheme can be compromised with complex model structures (≥ 15–20 state variables). Interestingly, our analysis showed a relatively higher change (central tendency shifts and standard deviation reductions) of the posterior moments of some parameters associated with the nutrient recycling in the system, i.e., dissolution and mineralization rates. Despite the aforementioned role of the model structure error terms and the high dimensional input space (35 model parameters) of the complex simulation model examined, some of the parameters representing feedback loops of the system played a somewhat more active role during the Bayesian updating process. Finally, the high coefficients of variation for the DO, DOC, and DSi error terms are indicative of the relatively low intra-annual variability characterizing these state variables (Arhonditsis *et al*. 2008a).

Aside from the probabilistic assessment of the water quality conditions, another benefit of the Bayesian parameter estimation is the alignment with the policy practice of adaptive management, i.e., an iterative implementation strategy that is recommended to address the often-substantial uncertainty associated with water quality model forecasts, and to avoid the implementation of inefficient and flawed management plans (Walters 1986). Adaptive implementation or "learning while doing" supports initial model forecasts of management schemes with post-implementation monitoring, i.e., the initial model forecast serves as the Bayesian prior, the post-implementation monitoring data serve as the sample information (the likelihood), and the resulting posterior probability (the integration of monitoring and modeling) provides the basis for revised management actions (Qian and Reckhow 2007). The probabilistic predictions for water quality variables of management interest (e.g., chlorophyll *a*, dissolved oxygen) can also be used to optimize water quality monitoring programs (Van Oijen *et al*. 2005). For example in Figure 8, the sections of the system where water quality conditions are more uncertain ("flat" distributions; C and D in the first map) should be more intensively monitored. These model predictions form the Bayesian prior which then is integrated (updated) with additional monitoring data to provide the posterior distribution. Based on the patterns of the posterior predictive distributions (where the predictive distribution for one site indicates a "high" probability of non-attaining water quality goals or, alternatively, an "unacceptably high" variance), we can determine again the optimal sampling design for water quality monitoring and assess the value of information (value of additional monitoring; *"Where should additional water quality data collection ef-*

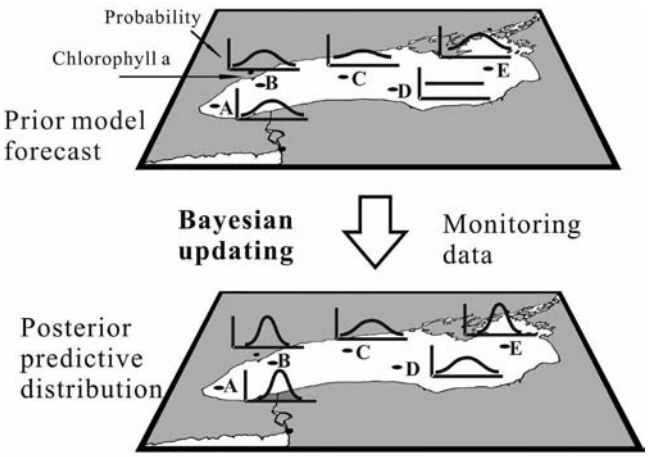

FIG. 8. Bayesian parameter estimation and optimization of the water quality monitoring using value of information concepts from decision theory.

forts be focused?"). The Bayesian inference and decision theory can also provide a coherent framework for decision making in problems of natural resources management (Dorazio and Johnson 2003). Management objectives can be evaluated by integrating the probability of use attainment for a given water quality goal with utility functions that reflect different socioeconomic costs and benefits. The water quality goals (resulting from specific management schemes) associated with the highest expected utility might then be chosen (Dorazio and Johnson 2003).

CONCLUSIONS

We illustrated a novel methodological framework that effectively addresses several aspects of model uncertainty (model structure, model parameters, initial conditions, and forcing functions) and explicitly examines how they can undermine the credibility of model predictions. We also demonstrated how the Bayesian parameter estimation can be used for assessing the exceedance frequency and confidence of compliance of different water quality criteria. The present analysis also highlighted the difficulty in unequivocally disaggregating the role of the uncertainty in model inputs and the error associated with the model structure (parameters versus model imperfection error terms); especially when using complex statistical formulations and models with multivariate outputs. Generally, our study provides overwhelming evidence that the coupling of the Bayesian calibration framework with complex overparameterized simulation models can negate the premise of attaining realistic forecasts while gaining mechanistic insights into the ecosystem dynamics. Thus, the use of complex models is advised only if existing prior information from the system can reasonably constrain the input parameter space, thereby ensuring model fit that is not founded on uninformative and/or fundamentally flawed ecological structures (e.g., unrealistic magnitudes of the various ecological processes). In cases where prior knowledge does not exist, it is advised to start with intermediate complexity models (4–10 state variables) and then gradually increase the complexity as more information becomes available (Arhonditsis *et al.* 2008b).

The latter assertions do not imply that this framework cannot accommodate the enormous complexity characterizing environmental systems, but rather are an indication that the rigid structure of complex mathematical models can be replaced by more flexible modeling tools (e.g., Bayesian networks) with the ability to integrate quantitative descriptions of ecological processes at multiple scales and in a variety of forms (intermediate complexity mathematical models, empirical equations, expert judgments), depending on available information (Borsuk *et al.* 2004). Regarding the spatial model resolution, our presentation was based on a single-compartment model for the sake of simplicity, but it should be acknowledged that the Bayesian framework can be easily employed with the segmentations of existing Great Lakes models, i.e., 5–10 completely-mixed boxes (Lam *et al.* 1987a, DiToro *et al.* 1987, Bierman *et al.* 2005). It is expected though that the use of finer grid resolutions will significantly increase the computation demands along with the simulation time required. To overcome this impediment, ongoing research should focus on the use of more flexible schemes, such as nested grid configurations that can reduce the computational time compared to the standard approach (one fixed grid size) and better capture the interplay between pollutant mixing/dispersion and food web dynamics in the nearshore areas, while the offshore water dynamics can be sufficiently reproduced with coarser spatiotemporal resolution. The patterns of the posterior uncertainty can then be used to further optimize the spatial model segmentation (e.g., splitting-up segments with flat posteriors or lumping segments with similar, narrow-shaped predictions) and avoid overly cumbersome modeling constructs that profoundly violate the parsimony principle.

Bearing in mind the pending reevaluation of the Great Lakes Water Quality Agreement, the Great Lakes community—as it did in the 1970s—has the opportunity to set the standard for the innovative use of mathematical models in support of decision-making. Despite the unresolved technical issues, we believe that the benefits from the Bayesian calibration scheme proposed, such as the assessment of uncertainty in model predictions and expression of model outputs as probability distributions, the alignment with the policy practice of adaptive management, and the optimization of the sampling design of monitoring programs can be particularly useful for stakeholders and policy makers when making decisions for sustainable environmental management in the Laurentian Great Lakes region.

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 Ontario Graduate Scholarships (Ontario Ministry of Training, Colleges and Universities). We wish to thank Craig Stow and Ram Yerubandi for their insightful comments on an earlier version of the manuscript. All the material pertinent to this study is available upon request from the second author.

GLOSSARY OF TERMS

Bayes' Theorem: is a theorem of probability theory originally stated by the Reverend Thomas Bayes. The theorem relates the conditional and marginal probability distributions of random variables, and tells how to update or revise beliefs in the light of new evidence from the study system.

Bayesian Inference: is a statistical approach in which all forms of uncertainty are expressed in terms of probabilities to represent the modification of our previous beliefs as a result of receiving new data. In the inference process, Bayes' Theorem is applied to obtain a posterior probability for a specific hypothesis, which considers both the prior probability and the observations from the study system.

Convergence: is the point in which MCMC sampling techniques eventually reach a stationary distribution. From this point on, the MCMC scheme moves around this distribution.

Credible Interval: is a posterior probability interval of a parameter or a model output. Credible intervals are the Bayesian counterparts of the confidence intervals used in frequentist statistics.

Likelihood Function: is a conditional function $[p(y|\theta)]$ considered as a function of its second argument (θ, model parameters) with its first argument (y, the data) held fixed. The likelihood function indicates how likely a particular population (model parameter set) can produce an observed sample.

Model Calibration: Calibration is the procedure by which the modeler attempts to find the best fit between computed and observed data by adjusting model parameters.

Markov Chain Monte Carlo (MCMC) Methods: are a class of algorithms for sampling from probability distributions based on the construction of a Markov chain that has the desired distribution as its stationary distribution. This procedure is used to generate a sequence of samples from a probability distribution that is difficult to be directly sampled.

Metropolis-Hastings Algorithm: is a rejection sampling algorithm, which generates a random walk using a proposal density and contains a method for rejecting proposed steps. It is one algorithm of Markov chain Monte Carlo methods.

Over Relaxation: At each MCMC iteration, a number of candidate samples are generated and one that is negatively correlated with the current value is selected. The time per iteration will be increased, but the within-chain correlations should be reduced and hence lower number of iteration may be necessary.

Posterior Distribution: is the conditional probability of a random event or an uncertain proposition that is derived when the relevant evidence from the study system is taken into account.

Prior Distribution: is a marginal probability that describes what is known about a variable in the absence of evidence from the study system.

Runge-Kutta Method: is a family of implicit and explicit iterative methods for the numerical approximation of solutions of ordinary differential equations.

Sensitivity Analysis: is the process by which the modeler attempts to evaluate the model sensitivity to the parameters selected, the forcing functions, or the state-variable submodels.

REFERENCES

Anderson, T.R. 2005. Plankton functional type modeling: running before we can walk? *J. Plankton Res.* 27(11):1073–1081.

_____. 2006. Confronting complexity: reply to Le Quere and Flynn. *J. Plankton Res.* 28(9):877–878.

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

_____, and Brett, M.T. 2005a. Eutrophication model for Lake Washington (USA) Part I—model description and sensitivity analysis. *Ecol. Model.* 187(2–3): 140–178.

_____, and Brett, M.T. 2005b. Eutrophication model for Lake Washington (USA) Part II—model calibration and system dynamics analysis. *Ecol. Model.* 187(2–3):179–200.

_____, Tsirtsis, G., and Karydis, M. 2002. The effects of episodic rainfall events to the dynamics of coastal marine ecosystems: applications to a semi-enclosed gulf in the Mediterranean. *Sea. J. Mar. Syst.* 35(3–4):183–205.

_____, Adams-VanHarn, B.A., Nielsen, L., Stow, C.A., and Reckhow, K.H. 2006. Evaluation of the current state of mechanistic aquatic biogeochemical modeling: citation analysis and future perspectives. *Environ. Sci. Technol.* 40(21):6547–6554.

_____, Qian, S.S., Stow, C.A., Lamon, E.C., and Reckhow, K.H. 2007. Eutrophication risk assessment using Bayesian calibration of process-based models: Application to a mesotrophic lake. *Ecol. Model.* 208(2–4):215–229.

_____, Papantou, D., Zhang, W., Perhar, G., Massos, E., and Shi, M. 2008a. Bayesian calibration of mechanistic aquatic biogeochemical models and benefits for environmental management. *J. Mar. Syst.* 73:8–30.

_____, Perhar, G., Zhang, W., Massos, E., Shi, M., and

Das, A. 2008b. Addressing equifinality and uncertainty in eutrophication models. *Water Resour. Res.* 44:W01420.

Beck, M.B. 1987. Water-quality modeling—a review of the analysis of uncertainty. *Water Resour. Res.* 23(11):1393–1442.

Beven, K.J. 1993. Prophecy, reality and uncertainty in distributed hydrological modeling. *Adv. Water Resour.* 16(1):41–51.

_____. 2001. *Rainfall-Runoff Modeling: The Primer*. New York: John Wiley.

_____. 2006. A manifesto for the equifinality thesis. *J. Hydrol.* 320(1–2):18–36.

Bierman, V.J., and Dolan, D.M. 1986. Modeling of phytoplankton in Saginaw Bay. 1. Calibration Phase. *J. Environ. Eng.-ASCE* 112(2):400–414.

_____, Kaur, J., DePinto, J.V., Feist, T.J., and Dilks, D.W. 2005. Modeling the role of zebra mussels in the proliferation of blue-green algae in Saginaw Bay, Lake Huron. *J. Great Lakes Res.* 31:32–55.

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.

_____, Stow, C.A., and Reckhow, K.H. 2004. A Bayesian network of eutrophication models for synthesis, prediction, and uncertainty analysis. *Ecol. Model.* 173(2–3):219–239.

Bowerman, W.W., Carey. J., Carpenter, D., Colborn, T., DeRosa, C., Fournier, M., Fox, G.A., Gibson, B.L., Gilbertson, M., Henshel, D., McMaster, S., and Upshur, R. 1999. Is it time for a Great Lakes Ecosystem Management Agreement separate from the Great Lakes Water Quality Agreement? *J. Great Lakes Res.* 25:237–238.

Brett, M.T., Arhonditsis, G.B., Mueller, S.E., Hartley, D.M., Frodge, J.D., and Funke, D.E. 2005. Non-point-source impacts on stream nutrient concentrations along a forest to urban gradient. *Environ. Manage.* 35(3):330–342.

Brooks, S.P., and Gelman, A. 1998. Alternative methods for monitoring convergence of iterative simulations. *J. Comput. Graph. Stat.* 7(4):434–455.

Cerco, C.F., and Cole, T.M. 1994. *CE-QUAL-ICM: a three-dimensional eutrophication model, version 1.0. User's Guide*. U.S. Army Corps of Engineers Waterways Experiments Station. Vicksburgh, MS.

Denman, K.L. 2003. Modeling planktonic ecosystems: parameterizing complexity. *Prog. Oceanogr.* 57 (3–4):429–452.

DiToro, D.M., Thomas, N.A., Herdendorf, C.E., Winfield, R.P., and Connolly, J.P. 1987. A post audit of a Lake Erie eutrophication model. *J. Great Lakes Res.* 13:801–825.

Dorazio, R.M., and Johnson, F.A. 2003. Bayesian inference and decision theory—A framework for decision making in natural resource management. *Ecol. Appl.* 13(2):556–563.

Edwards, A.M., and Yool, A. 2000. The role of higher predation in plankton population models. *J. Plankton Res.* 22(6):1085–1112.

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

Gilks, W.R., Richardson, S., and Spiegelhalter, D.J. (Eds.) 1998. *Markov Chain Monte Carlo in Practice*. London:Chapman & Hall/CRC.

Hartig, J.H., Zarull, M.A., and Law, N.L. 1998. An ecosystem approach to Great Lakes management: Practical steps. *J. Great Lakes Res.* 24:739–750.

Higdon, D., Kennedy, M., Cavendish, J.C., Cafeo, J.A., and Ryne, R.D. 2004. Combining field data and computer simulations for calibration and prediction. *SIAM J. Sci. Comput.* 26(2):448–466.

Jassby, A.D., and Platt, T. 1976. Mathematical formulation of relationship between photosynthesis and light for phytoplankton. *Limnol. Oceanogr.* 21(4):540–547.

Jorgensen, S.E., Nielsen, S.E., and Jorgensen, L.A. 1991. *Handbook of Ecological Parameters and Ecotoxicology*. New York:Elsevier Publications.

Krantzberg, G. 2004. Science must inform Great Lakes policy. *J. Great Lakes Res.* 30:573–574.

Lam, D.C.L., Schertzer, W.M., and Fraser, A.S. 1987a. Oxygen depletion in Lake Erie—modeling the physical, chemical, and biological interactions, 1972 and 1979. *J. Great Lakes Res.* 13:770–781.

_____, Schertzer, W.M., and Fraser, A.S. 1987b. A post-audit analysis of the NWRI 9-box water-quality model for Lake Erie. *J. Great Lakes Res.* 13:782–800.

Lampert, W., and Sommer, U. 1997. *Limnoecology: the ecology of lakes and streams*. New York:Oxford University Press.

Leon, L.K., Imberger, J., Smith, R.E.H., Hecky, R.E., Lam, D.C.L., and Schertzer, W.M. 2005. Modeling as a tool for nutrient management in Lake Erie: a hydrodynamics study. *J. Great Lakes Res.* 31(Suppl. 2): 309–318.

Ludwig, D. 1996. Uncertainty and the assessment of extinction probabilities. *Ecol. Appl.* 6(4):1067–1076.

McBride, G.B., and Ellis, J.C. 2001. Confidence of compliance: A Bayesian approach for percentile standards. *Water Res.* 35(5):1117–1124.

Mills, E.L., Casselman, J.M., Dermott, R.J., Fitzsimons, D., Gal, G., Holeck, K.T., Hoyle, J.A., Johannsson, O.E., Lantry, B.F., Makarewicz, J.C., Millard, E.S., Munawar, I.F., Munawar, M., O'Gorman, R., Owens, R.W., Rudstam, L.G., Schaner, T., and Stewart T.J. 2003. Lake Ontario: food web dynamics in a changing ecosystem (1970–2000). *Can. J. Fish. Aquat. Sci.* 60(4):471–490.

Minns, C.K., and Kelso, J.R.M. 2000. NO! It is time for a Great Lakes Ecosystem Management Agreement that SUBSUMES the Great Lakes Water Quality Agreement. *J. Great Lakes Res.* 26:1–2.

Neal, R. 1998. Suppressing random walks in Markov chain Monte Carlo using ordered over-relaxation. In *Learning in Graphical Models*, M.I. Jordan, ed., pp. 205–230. Dordrecht:Kluwer Academic Publishers.

Office of Water. 1997. *Guidelines for Preparation of the Comprehensive State Water Quality Assessments*. U.S. Washington, DC: Environmental Protection Agency.

Qian, S.S., and Reckhow, K.H. 2007. Combining model results and monitoring data for water quality assessment. *Environ. Sci. Technol*. 41(14):5008–5013.

Reichert, P., and Omlin, M. 1997. On the usefulness of overparameterized ecological models. *Ecol. Model*. 95(2–3):289–299.

Schindler, D.W. 2006. Recent advances in the understanding and management of eutrophication. *Limnol. Oceanogr*. 51(1):356–363.

Schladow, S.G., and Hamilton, D.P. 1997. Prediction of water quality in lakes and reservoirs. 2. Model calibration, sensitivity analysis and application. *Ecol. Model*. 96(3):111–123.

Spiegelhalter, D., Thomas, A., Best, N., and Lunn, D. 2003. *WinBUGS User Manual, Version 1.4*. Available at http://www.mrc-bsu.cam.ac.uk/bugs.

Steinberg, L.J., Reckhow, K.H., and Wolpert, R.L. 1997. Characterization of parameters in mechanistic models: a case study of a PCB fate and transport model. *Ecol. Model*. 97(1–2):35–46.

Stow, C.A., Reckhow, K.H., Qian, S.S., Lamon, E.C., Arhonditsis, G.B., Borsuk, M.E., and Seo, D. 2007. Approaches to evaluate water quality model parameter uncertainty for adaptive tmdl implementation. *J. Am. Water Resour. As*. 43(6):1499–1507.

Tian, R.C., Vezina, A.F., Starr, M., and Saucier, F. 2001. Seasonal dynamics of coastal ecosystems and export production at high latitudes: a modeling study. *Limnol. Oceanogr*. 46(8):1845–1859.

Van Oijen, M., Rougier, J., and Smith, R. 2005. Bayesian calibration of process-based forest models: bridging the gap between models and data. *Tree Physiol*. 25(7):915–927.

Walters, C.J. 1986. *Adaptive Management of Renewable Resources*. New York:McMillan.

Wild, P., Hordan, R., LePlay, A., and Vincent, R. 1996. Confidence intervals for probabilities of exceeding threshold limits with censored log-normal data. *Environmetrics*. 7(3):247–259.

Wroblewski, J.S. 1977. Model of phytoplankton plume formation during variable Oregon upwelling. *J. Mar*. 35(2):357–394.

Submitted: 26 January 2008
Accepted: 9 June 2008
Editorial handling: Joseph V. DePinto