ECOLOGICAL MODELLING 208 (2007) 215–229

available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/ecolmodel

Eutrophication risk assessment using Bayesian calibration of process-based models: Application to a mesotrophic lake

George B. Arhonditsis[a,*], Song S. Qian[b], Craig A. Stow[c], E. Conrad Lamon[b], Kenneth H. Reckhow[b]

[a] Department of Physical & Environmental Sciences, University of Toronto, Toronto, Ontario, Canada M1C 1A4
[b] Nicholas School of the Environment and Earth Sciences, Duke University, Durham, NC 27708, USA
[c] NOAA Great Lakes Environmental Research Laboratory, Ann Arbor, MI 48105-2945, USA

ARTICLE INFO

Article history:
Received 9 July 2006
Received in revised form
21 May 2007
Accepted 31 May 2007
Published on line 12 July 2007

Keywords:
Mechanistic models
Bayesian calibration
Environmental management
Risk assessment
Adaptive management implementation
Plankton dynamics
Eutrophication

ABSTRACT

We introduce the Bayesian calibration of process-based models to address the urgent need for robust modeling tools that can effectively support environmental management. The proposed framework aims to combine the advantageous features of both mechanistic and statistical approaches. Models that are based on mechanistic understanding yet remain within the bounds of data-based parameter estimation can accommodate rigorous and complete error analysis. The incorporation of mechanism improves the confidence in predictions made for a variety of conditions, while the statistical methods provide an empirical basis for parameter estimation and allow for estimates of predictive uncertainty. Our illustration focuses on eutrophication modeling but the proposed methodological framework can be easily transferred to a wide variety of disciplines (e.g., hydrology, ecotoxicology, air pollution). We examine the advantages of the Bayesian calibration using a four state variable (phosphate–detritus–phytoplankton–zooplankton) model and the mesotrophic Lake Washington (Washington State, USA) as a case study. Prior parameter distributions were formed on the basis of literature information, while Markov chain Monte Carlo simulations provided a convenient means for approximating the posterior parameter distributions. The model reproduces the key epilimnetic temporal patterns of the system and provides realistic estimates of predictive uncertainty for water quality variables of environmental interest. Finally, we highlight the benefits of Bayesian parameter estimation, such as the quantification of uncertainty in model predictions, optimization of the sampling design of monitoring programs using value of information concepts from decision theory, alignment with the policy practice of adaptive management, and expression of model outputs as probability distributions, that are perfectly suited for stakeholders and policy makers when making decisions for sustainable environmental management.

© 2007 Elsevier B.V. All rights reserved.

1. Introduction

In water quality assessment and management, mechanistic models are used to understand ecological processes, predict receiving aquatic ecosystem response to external nutrient loading changes, evaluate management alternatives, and support the policy making process (Reckhow and Chapra, 1999; Jorgensen and Bendoricchio, 2001). Although these modeling

* *Corresponding author*. Tel.: +1 416 208 4858; fax: +1 416 287 7279.
E-mail address: georgea@utsc.utoronto.ca (G.B. Arhonditsis).
0304-3800/$ – see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.ecolmodel.2007.05.020

constructs can be complex and contain much detail, their application involves substantial uncertainty contributed by both model structure and parameters. This uncertainty is not surprising because all models are drastic simplifications of reality that approximate the actual processes, i.e., essentially, all parameters are effective (spatially and temporally averaged) values unlikely to be represented by a fixed constant (Stow et al., 2003). Furthermore, ecological data are scarce or highly variable, so individual equations which are approximately correct in controlled laboratory environments may not collectively yield an accurate picture of ecosystem behaviour. Another problem that modellers do not seem to acknowledge is that the conventional model calibration, may provide the best fit of model input parameters to the dataset available at the moment, but it is specific to the given dataset at hand. As new data become available, the model should be recalibrated and in the common calibration practice there is no way of considering previous results. In this sense, we do not update previous knowledge about model input parameters, but rather we make the models dataset-specific (Kennedy and O'Hagan, 2001).

The conventional model calibration also does not address the well-known equifinality (poor identifiability) problem in which several distinct choices of model inputs lead to the same model outputs (many sets of parameters fit the model about equally well) (Beven and Binley, 1992). A main reason for the equifinality problem is that the causal mechanisms/hypotheses used for understanding how the system works internally is of substantially higher order than what can be externally observed (Beck, 1987). However, having a determination of model structure (and associated parameter values) that realistically reflects the natural system dynamics is particularly important when the model is intended for making predictions in the extrapolation domain (Arhonditsis and Brett, 2005a). For example, when a eutrophication model does not operate with realistic ecological structure (e.g., relative/absolute magnitudes of biological rates and transport processes), even if the fit between model outputs and observations is satisfactory, its credibility to provide predictions about how the system will respond under significantly different external nutrient loading conditions is very limited. In this case, the application of mechanistic models for extrapolative tasks is "an exercise in prophecy" rather than scientific action based on robust prognostic tools (Beven and Binley, 1992).

The importance of investigating the effects of uncertainty on model predictions has been extensively highlighted in the modeling literature (Reichert and Omlin, 1997; Omlin and Reichert, 1999). 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 and 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 analyses. The question of model uncertainty is important because models are used to identify polluters, direct the use of research dollars, and determine management strategies that have considerable social and economic implications (Clark et al., 2001). Erroneous model outputs and failure to account for uncertainty could provide misleading results and misallocation of the limited resources during the costly implementation of alternative environmental management schemes. For better model-based decision making, the uncertainty in model projections must be reduced or at least quantified and reported in a straightforward way that can be easily used by decision makers/policy planners (Reckhow, 1994; Arhonditsis et al., 2006).

The main objective of this study is to outline how environmental mechanistic (process-based) modeling can be integrated with Bayesian analysis, and the anticipated output will have broad applicability as a means for improving model forecasts and management actions over time and space. Our intent is to show that Bayesian calibration can be used to refine our knowledge of model input parameters, obtain insight into the degree of information the data contain about model inputs, and obtain predictions and uncertainty bounds for modeled output variables. Model uncertainty analysis essentially aims to quantify the joint probability distribution of model parameters and to make inference about this distribution; hence, the iterative nature of Bayes' Theorem is a convenient means to incorporate existing knowledge and update the joint distribution as new information becomes available. Bayesian parameter estimation is also a technique that is especially useful when scattered, multivariate information on the modeled system is available to improve the understanding of model parameters, but not enough to fully validate the mechanistic models. Bayesian inference should change the perspective of water quality modellers from seeking a single "optimal" value for each model parameter, to seeking a joint distribution of parameter sets, which then provide the basis for estimating model prediction error. Technically, we will show that there are better ways to parameterize mechanistic models; other than simply tuning (adjusting) model parameters until the modeller obtains "satisfactory" fit.

2. Methods

2.1. *Case study and model description*

Lake Washington was selected as a case study for testing the Bayesian calibration framework. Lake Washington is the second largest natural lake in Washington State, and is one of the best documented cases of successful restoration by sewage diversion (Edmondson, 1994). Currently, Lake Washington can be characterized as a mesotrophic ecosystem 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 dataset used for model calibration was based on a recent (1995–2001), spatially intensive (12 stations) limnological sampling program carried out by King County/Metro (http://dnr.metrokc.gov/wlr/waterres/lakes/LakeWashington.htm). Detailed description of the sampling network, the

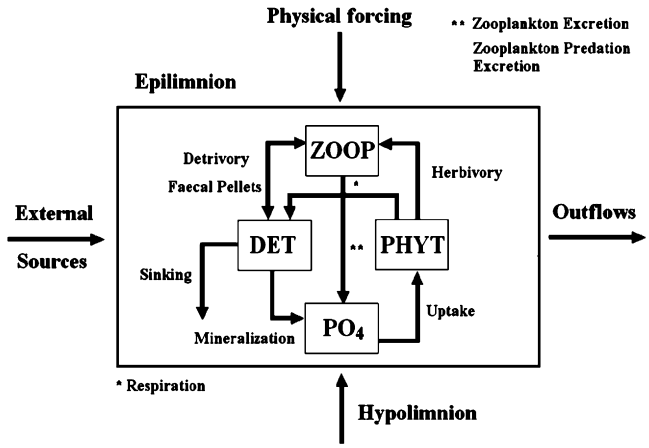

Fig. 1 – The phosphate–detritus–phytoplankton–zooplankton model used for reproducing the Lake Washington dynamics. Arrows indicate flows of matter through the system. System equations and parameter definitions are provided in Tables 1 and 2.

analytical methods used along with the data analysis is provided elsewhere (Arhonditsis et al., 2003; Arhonditsis and Brett, 2005b).

We opted for a parsimonious model structure that only considers the basic ecological processes underlying plankton dynamics in the mesotrophic environment of the Lake Washington epilimnion. The basic conceptual design of the model builds upon the results of a recent modeling study by Arhonditsis and Brett (2005a,b). We critically evaluated the outputs of the complex eutrophication model used to simulate multiple elemental cycles (org. C, N, P, Si, O), functional phytoplankton (diatoms, green algae and cyanobacteria) and zooplankton (copepods and cladocerans) groups, and developed a fairly simple single-compartment, four state variable approach (Fig. 1). Specifically, our single-compartment model considers the interplay (flows of matter) among the state variables: phosphate (PO_4), phytoplankton (PHYT), zooplankton (ZOOP), and detritus (DET); mathematically is described by the system of four ordinary differential equations presented in Table 1, while the definition of the model parameters is given in Table 2. The simulation model was solved numerically using the fourth-order Runge–Kutta method with a time step of 1 day. Some of the functional forms resemble those presented in the Edwards (2001) study, which also provided comprehensive examination of the model bifurcational behaviour. Here, we briefly describe the model structure along with the modifications introduced to simulate natural system dynamics (i.e., external forcing).

The governing equation for algal biomass considers phytoplankton production and losses due to basal metabolism, settling and herbivorous zooplankton grazing. Phosphorus limitation on phytoplankton growth follows the Michaelis–Menten kinetics, and the effects of the seasonal cycle of light and temperature are described by a trigonometric function $\sigma_{(t)}$. We used the same function of time introduced by Scheffer et al. (1997) to study periodically forced prey–predator systems, in which the amplitude and phase were adjusted to reproduce the mean physical conditions for the study period ($\varepsilon = 0.9$). Phytoplankton sinks out of the epilimnion at a constant rate, while the basal metabolism includes all internal processes that decrease algal biomass (respiration, excretion) as well as natural mortality. Zooplankton has two alternative food sources of equal palatability, i.e., phytoplankton and detritus. Both herbivory and detrivory were formulated using the Holling Type III function. A fraction of zooplankton grazing is assimilated and fuels growth, another fraction is recycled directly to phosphate, while the remaining fraction represents the faecal pellets and contributes to the detritus pool. Zooplankton losses encompass natural mortality and consumption by higher predators; we selected a sigmoid closure term that represents a "switchable"-type of predator behaviour controlled by a prey threshold concentration (equal to the half-saturation constant for predation; see Edwards and Yool, 2000). The effects of temperature on zooplankton metabolic activities were modeled by a similar trigonometric function assuming an approximate lag of 30 days. The specific parameterization for both zooplankton grazing and predation was adopted to more closely represent *Daphnia* dynamics, which is the dominant member of the Lake Washington zooplankton community (Arhonditsis and Brett, 2005a). A proportion of the zooplankton mortality/predation 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 variant diffusive mixing with the hypolimnion. Detritus sinks out of the epilimnion at a constant rate, while the watershed loadings (mainly representing the inflows of particulate phosphorus) were described

Table 1 – The specific functional forms of the eutrophication model

$$\frac{dPO_4}{dt} = -\frac{PO_4}{e+PO_4} a\sigma_{(t)}PHYT \times P/C_{phyto} + \frac{\beta\lambda((PHYT\times P/C_{phyto})^2+\omega DET^2)}{\mu^2+(PHYT\times P/C_{phyto})^2+\omega DET^2}\sigma_{(tz)}ZOOP \times P/C_{zoop} + \gamma\, d\sigma_{(tz)}\frac{ZOOP^3}{pred^2+ZOOP^2}P/C_{zoop} + \phi\sigma_{(t)}DET + k(1-\sigma_{(t)})(PO_{4(hypo)} - PO_4) + PO_{4exog} - outflows \times PO_4,\quad \sigma_{(t)} = \frac{1-\varepsilon\cos(2\pi t/365)}{1+\varepsilon}\sigma_{(tz)} = \frac{1-\varepsilon\cos(2\pi t/365-0.5)}{1+\varepsilon},\quad PO_{4(hypo)} = 11 + 3\sin\left(2\pi\left(\frac{t}{365}+0.3\right)\right),\quad PO_{4exog} = 0.052 + 0.02\sin\left(2\pi\left(\frac{t}{365}+0.12\right)\right),\quad outflows = 0.0028 + 0.0014\sin\left(2\pi\left(\frac{t}{365}+0.12\right)\right)$$

$$\frac{dPHYT}{dt} = \frac{PO_4}{e+PO_4} \times a\sigma_{(t)}PHYT - r\sigma_{(t)}PHYT - \frac{\lambda(PHYT\times P/C_{phyto})^2}{\mu^2+(PHYT\times P/C_{phyto})^2+\omega DET^2}\sigma_{(tz)}ZOOP - sPHYT - outflows \times PHYT$$

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

$$\frac{dDET}{dt} = r\sigma_{(t)}PHYT \times P/C_{phyto} + \frac{[(1-\alpha-\beta)(PHYT\times P/C_{phyto})^2-(\alpha+\beta)\omega DET^2]\lambda}{\mu^2+(PHYT\times P/C_{phyto})^2+\omega DET^2}\sigma_{(tz)}ZOOP \times P/C_{zoop} - \varphi\sigma_{(t)}DET - \psi DET + DET_{exog} - outflows \times DET,\quad DET_{exog} = 0.2 + 0.12\sin\left(2\pi\left(\frac{t}{365}+0.16\right)\right)$$

Table 2 – Parameter definitions of the eutrophication model

Parameter	Symbol	Units
Maximum phytoplankton (PHYT) growth rate	a^*	day^{-1}
Higher predation on zooplankton (ZOOP)	d^*	day^{-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^*	day^{-1}
Phytoplankton respiration rate	r^*	day^{-1}
Phytoplankton sinking loss rate	s^*	day^{-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_{zoo}	$0.029\,mg\,P\,(mg\,C)^{-1}$
Zooplankton growth efficiency	α^*	
Zooplankton excretion fraction	β^*	
Regeneration of zooplankton predation excretion	γ^*	
Maximum zooplankton grazing rate	λ^*	day^{-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^*	day^{-1}
Detritus sinking rate	ψ^*	day^{-1}

* Parameters used during the Bayesian calibration of the model.

by a sinusoidal function fitted to data from the local streams (Brett et al., 2005). Finally, similar sinusoidal functions were used for the remaining boundary conditions of the model, i.e., the exogenous dissolved phosphorus loadings, the hypolimnetic phosphate concentrations, and the outflows to the Lake Union Ship Canal (Arhonditsis and Brett, 2005a).

2.2. *Bayesian calibration*

Several recent studies have focused on the development of statistical methodologies that link mathematical models to the physical systems, i.e., frameworks that aim to provide probabilistic statements about the modeled systems using simulation models. Existing statistical methodological approaches to mathematical models addressed issues of uncertainty/sensitivity analysis (Saltelli et al., 2000), interpolation/emulation (Currin et al., 1991; Bates et al., 1996), calibration and prediction (Poole and Raftery, 2000; Kennedy and O'Hagan, 2001), and structural inadequacies of the available simulators (Craig et al., 2001; Goldstein and Rougier, 2004; Higdon et al., 2004). In this paper, we outline a conceptually similar statistical approach that aims 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. Our Bayesian framework explicitly accounts for the uncertainty in model inputs (model parameters and field data) and the discrepancy between the mathematical model and the environmental system. In this presentation, we place emphasis on the features of this methodology that demonstrate its appropriateness in the context of environmental modeling and management.

2.2.1. *The model is a perfect simulator of the environmental system (Model 1)*

The numerical solution of the eutrophication model is denoted as $f(\theta, x, y_0)$; where x 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 (i.e., the 14 parameters with asterisk in Table 2), and y_0 corresponds to the concentrations of the four state-variables at the initial time point t_0. At various settings for x, observations y for the four state variables are made from the environmental system and if we assume that the model perfectly describes its dynamics, then

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

where the ε_i's denote the observation (measurement) error and are usually assumed to be independent and identically distributed following a Gaussian distribution. Although the observed Lake Washington patterns provide evidence of a multiplicative measurement error (Arhonditsis et al., 2003; Fig. 2), we found that the required logarithmic transformation of the model outputs caused some computational problems. Therefore, we treated the errors additively and assumed the standard deviation to be proportional to the average monthly values for each state variable (Van Oijen et al., 2005). Specifically, we chose the monthly standard deviations to be 15% of the mean monthly values; a fraction that comprises both analytical error and interannual variability (Arhonditsis and Brett, 2005b).

Based on 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

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

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;

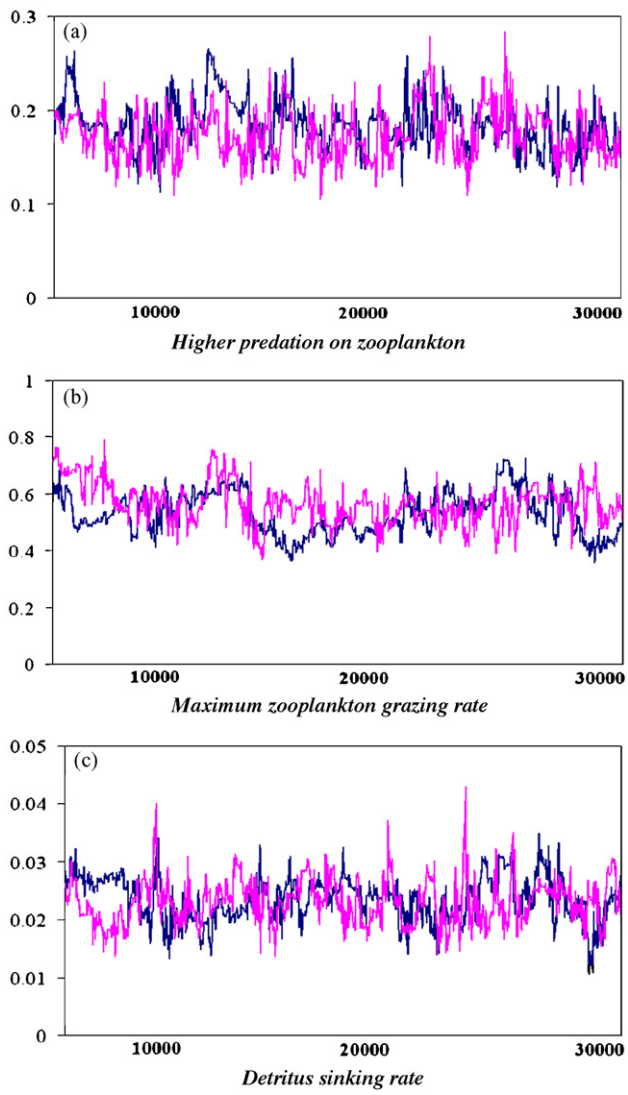

Fig. 2 – The Markov Chain Monte Carlo trace plots for the (a) higher predation on zooplankton, (b) maximum zooplankton grazing rate, and (c) detritus sinking rate (Model 2).

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

$$p(\theta, y_0|y) = \frac{p(y|f(\theta, x, y_0))p(\theta)p(y_0)}{\iint p(y|f(\theta, x, y_0))p(\theta)p(y_0)\,d\theta\,dy_0} \propto p(y|f(\theta, x, y_0))p(\theta)p(y_0) \quad (3)$$

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. The prior probability distribution, $p(\theta)$, reflects our knowledge on the relative plausibility of the different parameter values before calibration. In this study, the formulation of the prior density functions was based on field observations from the lake, laboratory studies, literature information, and expert judgment. Specifically, the characterization of the parameter distributions was similar to the protocol used in Steinberg et al. (1997), i.e., we identified the minimum and maximum values for each parameter and then we assigned lognormal distributions parameterized such that 95% of their values were lying within the identified ranges. 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 20% of the mean value for each state variable j; a fraction that comprises both analytical error, interannual variability, 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

$$\begin{aligned} p(\theta, y_0|y) \propto & \prod_{j=1}^{m} (2\pi)^{-n/2} |\Sigma_{\varepsilon j}|^{-1/2} \\ & \times \exp\left[-\frac{1}{2}[y_j - f_j(\theta, x, y_0)]^T \Sigma_{\varepsilon j}^{-1} [y_j - f_j(\theta, x, y_0)]\right] \\ & \times (2\pi)^{-l/2} |\Sigma_\theta|^{-1/2} \prod_{k=1}^{l} \frac{1}{\theta_k} \\ & \times \exp\left[-\frac{1}{2}[\log\theta - \theta_0]^T \Sigma_\theta^{-1} [\log\theta - \theta_0]\right] \\ & \times (2\pi)^{-m/2} |\Sigma_{y0}|^{-1/2} \\ & \times \exp\left[-\frac{1}{2}[y_0 - y_{0m}]^T \Sigma_{y0}^{-1} [y_0 - y_{0m}]\right] \end{aligned} \quad (4)$$

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 (1995–2001); and $\Sigma_{y0} = I_m(0.20)^2 \cdot y_{0m}^T \cdot y_{0m}$.

2.2.2. *The model is an imperfect simulator of the environmental system (Model 2)*

An augmentation of our statistical formulation explicitly recognizes that the model imperfectly represents the dynamics of the environmental system (e.g., missing key ecological processes, erroneous formulations, misspecified forcing functions). In this case, an observation i for the state variables j, y_{ij}, can be described as

$$y_{ij} = f(\theta, x_i, y_0) + \delta_j + \varepsilon_{ij}, \quad i = 1, 2, 3, \ldots n \text{ and } j = 1, \ldots, m \quad (5)$$

where the stochastic term δ_j accounts for the discrepancy between the model $f(\theta, x, y_0)$ and the natural system. In this study, we assumed that the model discrepancy is 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), which is slightly differ-

ent from the statistical formulation introduced by Higdon et al. (2004 see Discussion). With this assumption, the likelihood function will be

$$p(y|f(\theta, x, y_0)) = \prod_{j=1}^{m}(2\pi)^{-n/2}|\Sigma_{Tj}|^{-1/2} \times \exp\left[-\frac{1}{2}[y_j - f_j(\theta, x, y_0)]^T \Sigma_{Tj}^{-1}[y_j - f_j(\theta, x, y_0)]\right] \quad (6)$$

$$\Sigma_{Tj} = \Sigma_{\delta j} + \Sigma_{\varepsilon j} \quad (7)$$

where $\Sigma_{\delta j} = I_n \cdot \sigma_j^2$ corresponds to the additional stochastic term of Model 2; and the prior densities $p(\sigma_j^2)$ 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:

$$\begin{aligned} p(\theta, y_0, \sigma^2|y) \propto & \prod_{j=1}^{m}(2\pi)^{-n/2}|\Sigma_{Tj}|^{-1/2} \\ & \times \exp\left[-\frac{1}{2}[y_j - f_j(\theta, x, y_0)]^T \Sigma_{Tj}{}^{-1}[y_j - f_j(\theta, x, y_0)]\right] \\ & \times (2\pi)^{-l/2}|\Sigma_\theta|^{-1/2}\prod_{k=1}^{l}\frac{1}{\theta_k} \\ & \times \exp\left[-\frac{1}{2}[\log\theta - \theta_0]^T \Sigma_\theta{}^{-1}[\log\theta - \theta_0]\right] \\ & \times (2\pi)^{-m/2}|\Sigma_{y0}|^{-1/2} \\ & \times \exp\left[-\frac{1}{2}[y_0 - y_{0m}]^T \Sigma_{y0}{}^{-1}[y_0 - y_{0m}]\right] \\ & \times \prod_{j=1}^{m}\frac{1}{up_j - lo_j} \end{aligned} \quad (8)$$

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

2.2.3. *Numerical approximations for posterior distributions*

Sequence of realizations from the posterior distribution of the two models were obtained using Markov chain Monte Carlo (MCMC) simulations (Gilks et al., 1998). Specifically, we used the general normal-proposal Metropolis algorithm as it is implemented in the WinBUGS software (Spiegelhalter et al., 2003); 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. Based on the Steinberg et al. (1996) findings, the posterior simulations were based on multiple chains from starting points dispersed around the parameter space. We found that some of the initial parameter vectors resulted in unstable solutions, i.e., solutions of the dynamical system that tended to infinity. 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) 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 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.

2.2.4. *Assessment of the goodness-of-fit and model comparisons*

Assessment of the goodness-of-fit between the model predictions and the observed data was based on the posterior predictive p-value, i.e., 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; while 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 χ^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, i.e., 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

$$B_{12} = \frac{pr(y|M_1)}{pr(y|M_2)} \quad (9)$$

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

$$pr(y|M_k) = \int pr(y|M_k, \Theta_k)\pi(\Theta_k|M_k)\,d\Theta_k \quad (10)$$

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

$$\overline{pr(y|M_k)} = \left\{\frac{1}{m}\sum_{i=1}^{m} pr(y|M_k, \Theta_k^{(i)})^{-1}\right\}^{-1} \quad (11)$$

the harmonic mean of the likelihood values (Kass and Raftery, 1995).

3. Results

The MCMC trace plots for three model parameters (higher predation on zooplankton, maximum zooplankton grazing rate, detritus sinking rate), after discarding the "burn-in" (i.e., the period of adjustment to the appropriate conditional distribution), are shown in Fig. 2. The irregular patterns of these sequences of iterations are characteristic of the MCMC simulation, because the aim is to create a Markov process whose stationary distribution approximates the joint posterior distribution of all the stochastic nodes of the model rather than identify the global optimum for maximizing the model fit (Gelman et al., 1995). Generally, we noticed that the sequences converged rapidly (≈5000 iterations), while the statistics reported in this study were based on the last 25,000 draws by keeping every 4th iteration (i.e., an appropriate thinning to avoid serious autocorrelation problems). The uncertainty underlying the values of the 14 model parameters before and after the Bayesian calibration is depicted on the respective prior and marginal posterior distributions (Table 3 and Fig. 3). Generally, the coefficients of variation (or "relative standard deviation"; CV) of the posterior parameter distributions were significantly reduced when using the first statistical formulation (Model 1); characteristic examples were the detritus sinking rate (ψ), the phytoplankton sinking loss rate (s), the assimilation efficiency (a), and the maximum growth rate (a) with a decrease from 50 to 90%. On the other hand, the inclusion of the stochastic term that accounts for the discrepancy between the model and the real system resulted in significantly higher CV parameter values, and – in some cases – our knowledge did not improve relative to what we knew prior to the calibration (e.g., higher predation on zooplankton, maximum zooplankton grazing rate, maximum growth rate, cross-thermocline exchange rate, and zooplankton excretion fraction). The latter finding indicates that the discrepancy term mainly improves our knowledge on the natural system dynamics (see next paragraph) but gives little information regarding the values of the calibration vector. In a general context, however, the Higdon et al. (2004) study showed 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 (see the results reported in their Figs. 3 and 11). It should also be noted that the central tendency of the updated distributions of the phytoplankton respiration rate (r) and zooplankton grazing half-saturation coefficient (μ) with the Model 1 was significantly higher than the prior assigned values. To avoid these major shifts on the two posterior distributions, we used truncated prior distributions but the subsequent calibration resulted in inferior results comparing with those found from the full parameter space.

The Bayesian calibration can also be used to determine the correlation structure among the model parameters using the MCMC posterior samples (Table 4). Some of these relationships have plausible physical explanation that can be derived from the model equations. If the phytoplankton respiration rate (r) is high, for example, the detritus remineralization rate (φ) will also be high so as the amount of detritus (or particulate phosphorus) in the water column not be greater than that observed. Likewise, if the maximum phytoplankton growth rate (a) is high, then the half-saturation constant for PO_4 uptake (e) should also be high in order to accurately represent the observed epilimnetic phytoplankton dynamics. There were

Table 3 – Prior and Markov Chain Monte Carlo posterior estimates of the mean values and standard deviations of the model stochastic nodes

Parameter	Prior		Model 1		Model 2	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
a	1.801	0.515	1.605	0.133	1.153	0.302
d	0.164	0.026	0.186	0.017	0.199	0.028
pred	36.49	15.64	28.22	4.693	34.96	9.404
e	12.16	5.212	6.343	1.686	8.215	4.579
k	0.036	0.010	0.020	0.003	0.028	0.007
r	0.136	0.066	0.378	0.082	0.227	0.069
s	0.080	0.085	0.054	0.006	0.061	0.012
a	0.360	0.103	0.459	0.032	0.362	0.050
β	0.431	0.077	0.403	0.052	0.439	0.071
γ	0.431	0.077	0.421	0.072	0.433	0.079
λ	0.535	0.077	0.664	0.035	0.568	0.091
μ	5.055	1.285	9.516	0.781	5.413	1.810
φ	0.155	0.060	0.089	0.026	0.072	0.018
ψ	0.113	0.072	0.026	0.002	0.024	0.004
$PO_{4(0)}$	13.57	2.714	13.08	1.661	13.48	2.124
$PHYT_{(0)}$	61.94	12.39	69.35	10.90	67.69	16.94
$ZOOP_{(0)}$	15.97	3.194	15.35	1.487	19.77	5.040
$DET_{(0)}$	15.68	3.136	18.80	2.636	17.20	3.387
σ_{PO_4}					1.537	0.578
σ_{chla}					113.9	40.92
σ_{zoop}					22.99	9.456
σ_{det}					0.504	0.815

also some relationships that seem counterintuitive (based on the model setup), such as the positive correlation between the zooplankton excretion fraction (β) and detritus sinking rate (ψ). Given that higher values for both parameters normally result in lower detritus levels in the system, a negative relationship under which the two terms cancel each other out would have seemed more plausible. The positive correlation probably reflects the predominance of other ecological paths considered in our eutrophication model, e.g., higher fraction of the zooplankton excretion provides phosphate which fuels phytoplankton growth, and then the resulting increase in phytoplankton respiration and zooplankton grazing supply the detritus pool. Thus, higher detritus sinking rates probably compensate for this increase, and the model can still provide a reasonable fit to the observed data. The latter pattern is particularly pronounced during the summer stratified period when the system is strongly phosphorus limited and the phytoplankton growth mainly depends on the internal loading (see following discussion). Furthermore, the implementation of principal component analysis on the resulting correlation matrix provides evidence of several distinct clusters that characterize the posterior parameter variability (Legendre and Legendre, 1998). Based on the rotated (normalized varimax) component loadings (Fig. 4), we can distinguish one cluster that consists of the respiration (r) and detritus remineralization (φ) rates; a second cluster that comprises the half-saturation constant for grazing (μ), the maximum phytoplankton growth rate (a), and the half-saturation constant

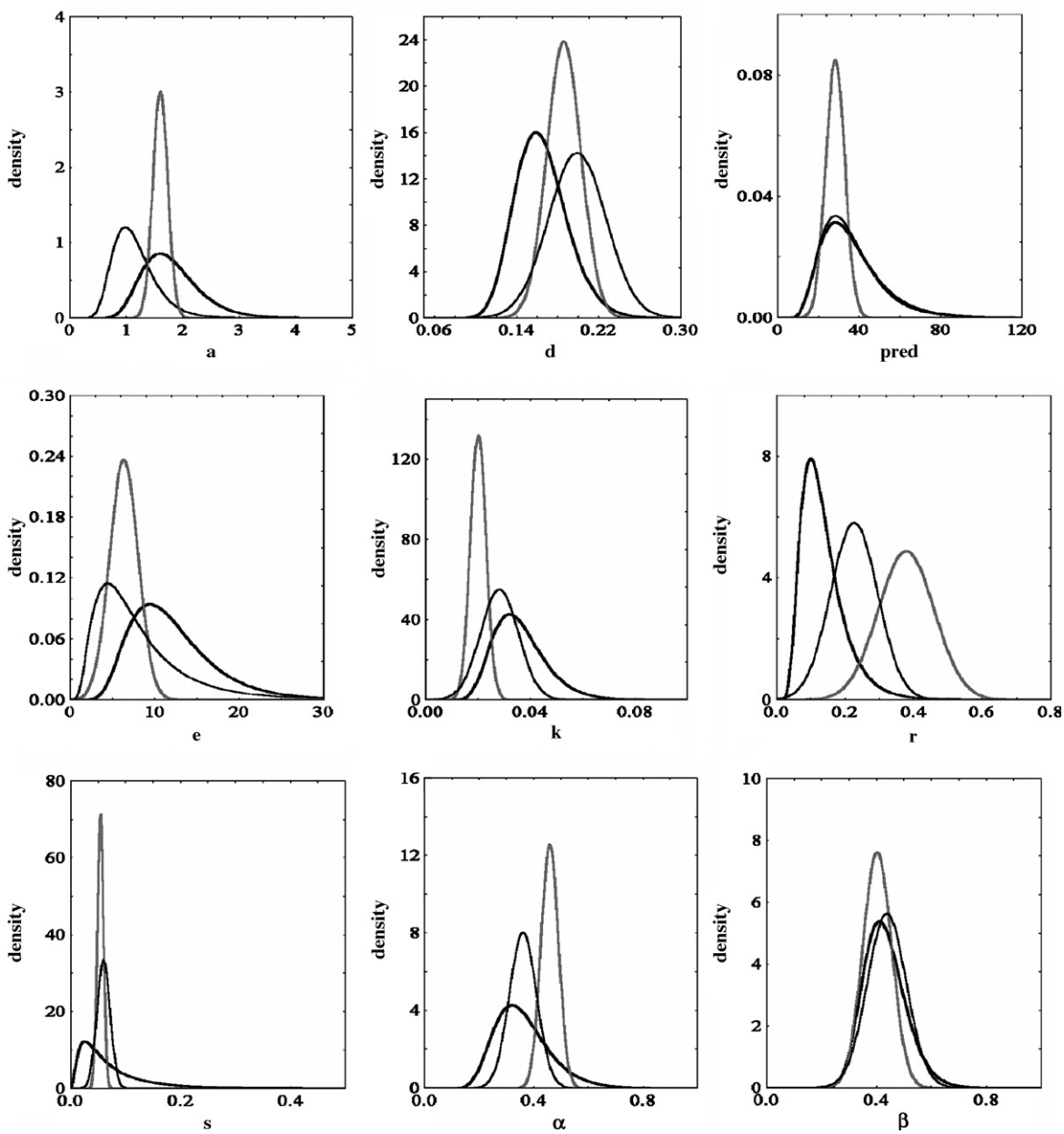

Fig. 3 – Prior (thick black lines) and posterior (Model 1: gray lines and Model 2: thin black lines) distributions of the eutrophication model parameters. The posteriors depict smoothed kernel density estimates based on 12,500 MCMC samples from the two models.

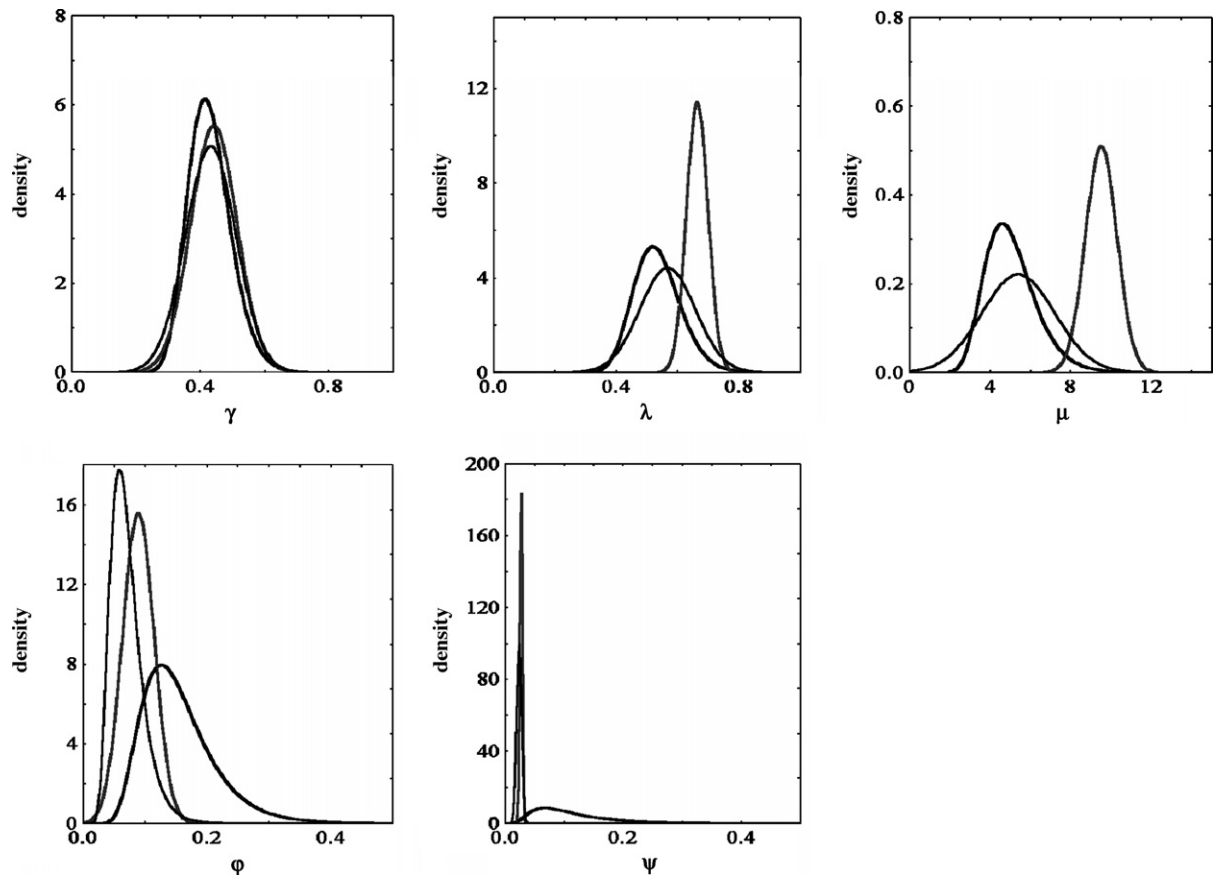

Fig. 3 – (*Continued*).

for PO_4 uptake (e); a third group that includes the phytoplankton sinking rate (s), the zooplankton growth efficiency (α) and higher predation (d); while the remaining parameters, i.e., maximum grazing rate (λ), the zooplankton excretion fraction (β), the cross-thermocline exchange rate (k), the detritus (ψ) sinking rate, the regeneration of zooplankton predation excretion (γ) and half-saturation constant for predation (pred) were not classified in a coherent cluster. Finally, we found a similar (to the parameters) pattern between the two statistical formulations in terms of the CV values of the prior and posterior distributions of the initial conditions; the first model was more informative and resulted in significantly reduced CV values (Table 3 and Fig. 5). We also note the relatively high CV values of the error terms (σ_j) that account for the discrepancy between the simulation model and the natural system (Model 2), especially for detritus (>100%).

Table 4 – Correlation matrix of the eutrophication model parameters

	a	α	β	d	e	γ	k	λ	μ	φ	pred	ψ	r	s
a														
α	0.027													
β	0.278	**−0.719**												
d	−0.018	−0.205	0.248											
e	**0.746**	−0.208	0.393	0.091										
γ	−0.027	0.122	−0.300	−0.131	0.042									
k	0.204	0.026	−0.047	−0.115	0.198	0.049								
λ	0.244	−0.109	0.273	**0.522**	0.456	0.033	0.030							
μ	0.171	0.467	−0.249	−0.325	0.207	0.208	0.125	0.409						
φ	−0.306	0.044	−0.164	−0.374	**−0.556**	−0.080	−0.036	**−0.683**	−0.146					
pred	−0.178	−0.265	0.063	−0.215	−0.091	−0.066	0.064	−0.191	0.234	0.463				
ψ	0.120	−0.753	**0.579**	0.010	0.219	−0.080	0.074	0.038	−0.341	0.028	0.341			
r	−0.178	0.007	−0.154	−0.351	**−0.586**	−0.078	−0.093	**−0.640**	−0.204	**0.776**	0.229	0.102		
s	0.260	**0.543**	−0.158	−0.002	−0.098	−0.089	0.095	0.123	0.143	−0.148	−0.413	−0.469	0.000	

Bold numbers correspond to correlation coefficients with absolute value greater than 0.5.

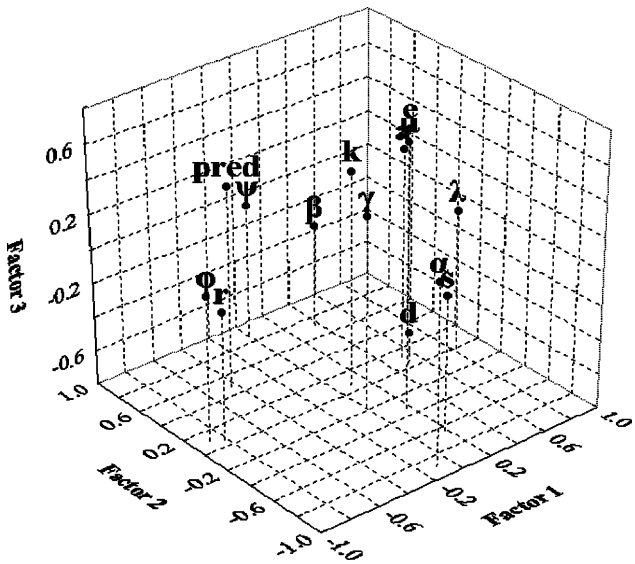

Fig. 4 – Principal component analysis of the 14 model parameters based on Markov chain Monte Carlo posterior samples from Model 2. [Significant principal components were rotated using the normalized varimax strategy (raw factor loadings divided by the square roots of the respective communalities), and the three PCs explained 89% of the observed variability.]

The two models were favourably supported by the data and were accepted on the basis of their posterior predictive *p*-values (0.124 and 0.385 for Models 1 and 2, respectively). The Bayes factor value $B_{21} = 2.85$ did not provide strong evidence in support of one of the two alternative models but did reflect a better predictive performance of the Model 2. The comparison between the observed (monthly averages over the 7-year period, 1995–2001) and posterior predictive monthly distributions for the four state variables illustrates some features of the Bayesian calibration (Fig. 6). The Model 1 provides a moderate fit to the four state variables; 83.3% (10 of 12) of the phosphate, 58.3% (7 of 12) of the total phosphorus, 83.3% (10 of 12) of the chlorophyll concentrations and 83.3% (10 of 12) of the zooplankton biomass data were included within the 95% credible intervals. On the other hand, the addition of the discrepancy term in the second statistical formulation (Model 2) has significantly improved the results and all the observed monthly values were included within the 95% credible intervals. Regarding the central tendency of the predictive monthly distributions, we found that the predicted median values were lower than the observed spring phosphorus levels (Model 1) and also underestimated the spring maximum phytoplankton and zooplankton biomass. [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 provided fairly similar results and only slightly improved the representations of the mid/late spring plankton dynamics.] These results can partly be attributed to the inaccurate representation of the physical conditions (weather 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. 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 mean of a Gaussian distribution with standard deviation assumed to be 10% of the mean values). The predicted median phosphate–plankton biomass values were closer to the lake seasonal dynamics but, not surprisingly, were also accompanied by wider prediction bands (not presented here).

The estimation of the exceedance frequency for different chlorophyll levels can be based on the respective marginal/cumulative predictive distribution (Fig. 7), e.g., there is 15% probability that chlorophyll will be higher than 4 μg/L during the summer stratified period in Lake Washington. Furthermore, the joint predictive distributions of total phosphorus, chlorophyll, and zooplankton abundance can offer insights into the relationships among the limiting nutrient, primary producers and herbivory over the same period. For example, the model predictions suggest that if our target for the chlorophyll level of 4 μg/L is an exceedance frequency lower than 10%, then the summer TP concentrations should be lower than 14 μg/L (Fig. 8a). The distinctly positive relationship between the summer phytoplankton community and zooplankton biomass provide evidence for co-dependence and tight primary producer–herbivore association, especially when the zooplankton biomass lies within the range of 60–80 μg/L (Fig. 8b). This result can be explained by the findings of earlier experimental (Richey, 1979) and modeling (Arhonditsis and Brett, 2005b) studies which estimated that zooplankton nutrient recycling (mostly by *Daphnia pulicaria* and *Daphnia thorata*) provided 60–90% of the phosphorus supply to the mixed layer during the summer stratified period. Thus, zooplankton nutrient recycling fuels phytoplankton growth, which in turn has a positive feedback and sustains herbivore biomass. The latter presentation of the model outputs as probabilistic assessment of water quality along with the underlying ecological interactions conveys significantly more information than point (single-valued) estimates in regards to uncertainty and can make the model results more credible and appealing to decision makers and stakeholders. Finally, the state-space for total phosphorus–phytoplankton–zooplankton dynamics, based on the implementation of a spline fitting on the MCMC posterior samples (Chapra and Canale, 1998), provides a comprehensive presentation of the combined bottom-up and top-down effects on phytoplankton during the summer stratified period in Lake Washington (Fig. 9).

4. Discussion

The general lack of uncertainty estimates for most environmental models, the arbitrary selection of higher – and often unattainable – threshold values for environmental variables (quality goals/standards) as a hedge against unknown forecast errors, risky model-based management decisions and unanticipated system responses are the norm in current management practice. The ubiquitous and often substantial model

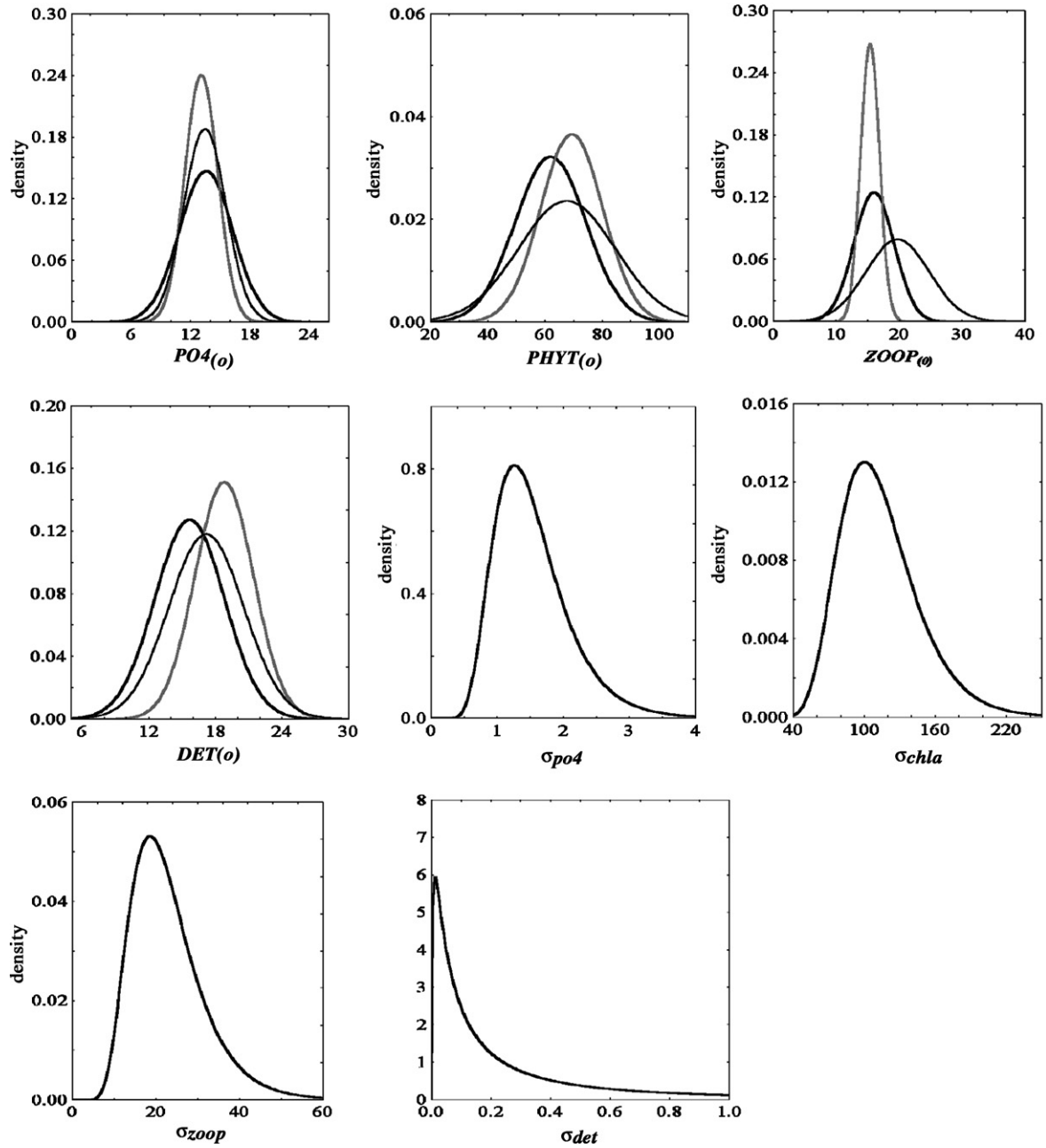

Fig. 5 – Prior (thick black lines) and posterior (Model 1: gray lines and Model 2: thin black lines) distributions of the initial conditions and the error terms (i.e., discrepancy between the simulation model and the natural system) for the four state variables.

uncertainty is a key reason why adaptive implementation is often considered the only defensible strategy in environmental management (Walters, 1986). If, based on current science, the initial management plan is likely to be inefficient or inadequate, why not design the management implementation process so that it accommodates error and correction? Yet, while the concept of adaptive implementation makes sense as a pragmatic solution to scientific uncertainty, practical matters related to technical approaches and policy issues remain (Borsuk et al., 2002). We presented an illustrative example of a methodological tool (Bayesian calibration of mechanistic models) that can be easily engaged with the policy practice of adaptive management and address some of the technical difficulties. The proposed framework combines the advantageous features of both mechanistic and statistical approaches. Models with mechanistic foundation that are combined with a statistical configuration; the former component can provide more reliable predictions of system behaviour, while the latter allows for an empirical parameter estimation and rigorous assessment of predictive uncertainty. We used a Bayesian formulation that can simultaneously accomplish several goals, i.e., offer insights into the degree of information the data contain about model inputs, quantify the dependence structure (correlation) among parameter estimates, and obtain pre-

dictions along with uncertainty bounds for modeled output variables.

Some of the technical advances and benefits for environment management from the Bayesian calibration of process-based models are:

i) *Identification problem*: As it was indicated in the present study, prior knowledge of the magnitudes of ecological processes can be converted into probability distributions that reflect the relative plausibility of different values of the respective model parameters, which then can be included into the "prior–likelihood–posterior" update cycles. Viewing model calibration as an inverse problem, this practice increases the likelihood of model solutions that more realistically reflect the internal structure of the system and avoid getting "good results for the wrong reasons". Furthermore, the Bayesian inference techniques offer an effective strategy to overcome the typical problem

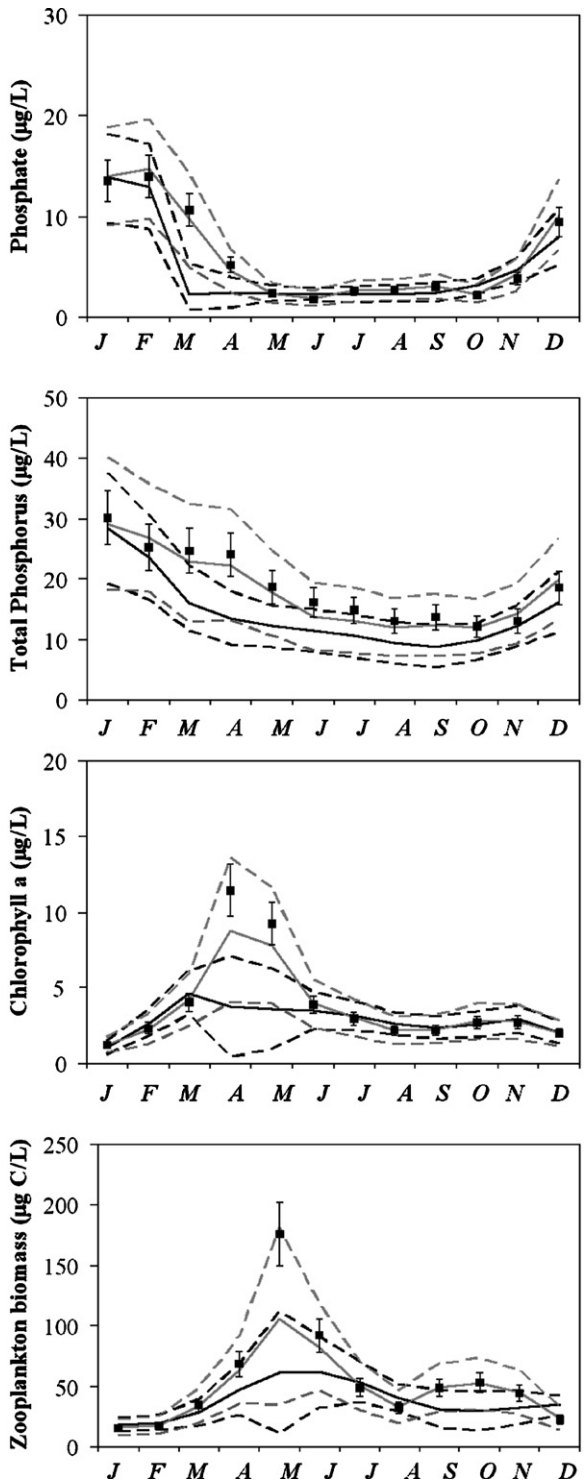

Fig. 6 – Comparison between the observed and posterior predictive monthly distributions for phosphate, total phosphorus, chlorophyll *a* (1 g carbon = 20 mg chlorophyll) and total zooplankton abundance for the Lake Washington eutrophication model (Model 1: black line and Model 2: gray line). Single dots 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 7-year period, 1995–2001. Dashed lines correspond to the 2.5 and 97.5% credible intervals. The modeled total phosphorus concentrations comprise three phosphorus pools; i.e., phosphate, detritus and phosphorus sequestered in phytoplankton cells.

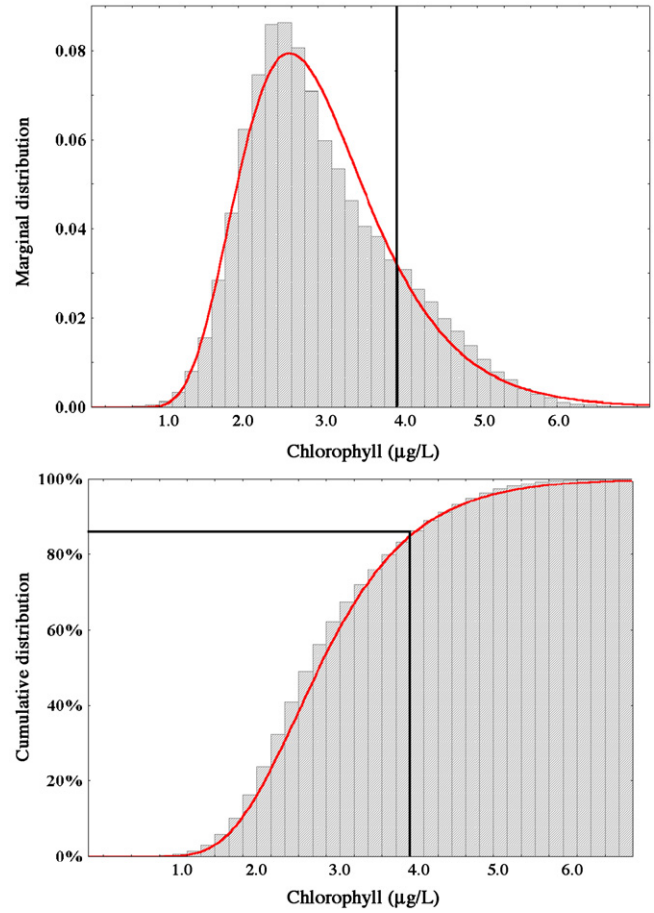

Fig. 7 – Marginal and cumulative predictive distribution of chlorophyll during the summer stratified period in Lake Washington. [Results based on 12,500 Markov chain Monte Carlo posterior samples from Model 2.]

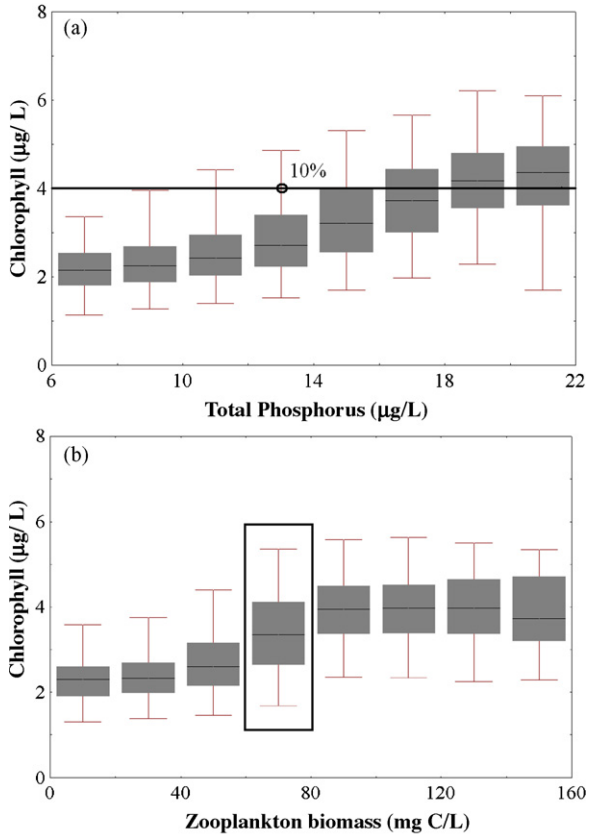

Fig. 8 – Predictive chlorophyll distributions (median values and 95% credible sets) for different (a) total phosphorus and (b) zooplankton abundance levels during the summer stratified period in Lake Washington (Model 2).

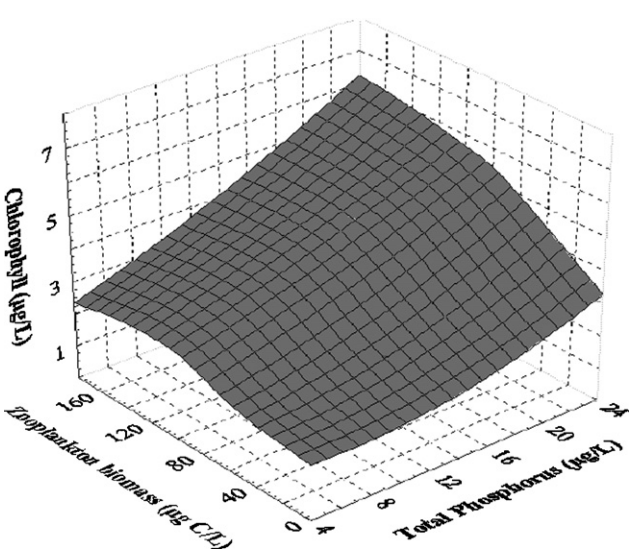

Fig. 9 – State-space for total phosphorus–phytoplankton–zooplankton dynamics during the summer stratified period in Lake Washington. [Spline fitting was implemented on the 12,500 Markov chain Monte Carlo posterior samples from Model 2.]

of identification in environmental models, because the use of additional information (along with the calibration dataset) reduces the disparity between what ideally we want to learn (internal description of the system) and what can realistically be observed; the primary reason for poor model identifiability (Beck, 1987; Omlin and Reichert, 1999).

ii) *Realistic predictive uncertainty estimates*: The Bayesian approach generates a posterior predictive distribution that takes into account both the uncertainty about the parameters and the uncertainty that remains when the parameters are known (Kennedy and O'Hagan, 2001). As a result, the predictive uncertainty estimates from the Bayesian inference techniques are more realistic (usually larger than those based on the classical procedures), and thus more appropriate for environmental management (Reichert and Omlin, 1997). For example, as it was shown in our study (Figs. 6–8), the – often misleading – deterministic statements can be avoided and the water quality goals can be set by explicitly acknowledging an inevitable risk of non-attainment, the level of which is subject to decisions that reflect different socioeconomic values and environmental concerns.

iii) *Adaptive management implementation and optimization of monitoring programs*: The Bayesian (iterative) nature of the proposed approach is conceptually similar to 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 avoid implementation of inefficient and flawed management plans. Adaptive implementation or "learning while doing" augments initial model forecasts of management schemes with post-implementation monitoring; 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 (and improved) management actions. The probabilistic assessment for water quality variables of management interest (e.g., chlorophyll *a*) that incorporates all possible sources of uncertainty (model inadequacy, parameter uncertainty, observation error) can also indicate where the limited monitoring resources should be focused (Van Oijen et al., 2005). Based on the patterns of the posterior predictive distributions (where the predictive distribution for one site indicates a "low" probability of 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 efforts be focused?*").

iv) *Bayesian model averaging*: Although beyond the scope of the present paper, a logical next step of the proposed framework would be to examine how the uncertainty patterns of the model predictions and the posterior parameter distributions can be used to optimize model complexity by omitting highly uncertain or including missing ecological processes. An alternative approach would be to recognize that there is no true model of an

ecological system, but rather several adequate descriptions of different conceptual basis and structure (Reichert and Omlin, 1997). Therefore, rather than picking the single "best-fit" model to predict future system responses, we can use Bayesian model averaging to provide a weighted average of the forecasts from different models (Hoeting et al., 1999; Stow et al., 2004). For example, an appealing future next step of this methodological framework will be the combination of the present fairly simple (4 state variable) model with the existing complex eutrophication (21 state variable) model for Lake Washington (Arhonditsis and Brett, 2005a,b). Hence, by acknowledging that models at both ends of the complexity spectrum have different strengths and weaknesses, Bayesian model averaging can be used to provide better average predictive ability and overcome the ambiguity regarding model selection or the risk of basing ecological forecasts on a single model.

One of the major disadvantages of the Bayesian inference techniques that has prohibited their broader application in the common modeling practice has been their high computation demands (Reichert and Omlin, 1997; Reichert et al., 2002). In our study, we showed that the MCMC procedure can overcome the lack of analytical expressions for the posterior distribution; a typical problem with the nonlinear parameterizations used in eutrophication modeling. 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 predicted model outputs. Although more advanced procedures are available (Gilks et al., 1998; Robert and Casella, 1999), several modeling studies from a variety of disciplines indicated that even simpler MCMC schemes can effectively sample high dimensional parameter spaces and multivariate outputs (Hegstad and More, 2001; Lee et al., 2002; Van Oijen et al., 2005). Alternatively, Higdon et al. (2004) proposes a compromise between the "fidelity of the simulator" and the "simulation speed" claiming that a comprehensive examination of the posterior distribution of a simple model can be more informative than an insufficient posterior approximation of a more complex model (see their charged particle accelerator example). Finally, it should be noted that the additional model complexity does not necessarily imply more MCMC runs; if the number of parameters that drive the model outputs does not change, then the number of runs required to sufficiently approximate the posterior will be approximately the same (Jansen and Hagenaars, 2004).

In the present study, our aim was to illustrate how some features of the Bayesian model calibration can assist environmental management, and thus our presentation was neither exhaustive in terms of the technical aspects of this framework nor informative regarding its ability to produce extrapolated predictions, i.e., reproduce ecological patterns under significantly different external conditions such as increased nutrient loading. For example, a follow-up study demonstrates how the Bayesian approach supports management decisions by assessing both the exceedance frequency and confidence of compliance of different water quality criteria (Zhang and Arhonditsis, submitted). Furthermore, while the previously mentioned stochastic treatment of the forcing functions can also be used to accommodate the role of interannual variability under the current management practices, the extrapolation in new, significantly different, regions of system behaviour requires special consideration [see Bayarri et al. (2002) for some basic principles underlying the model extrapolation task]. Finally, an interesting next step from a methodological point of view will be the relaxation of the assumption that the model discrepancy is invariant with the input conditions (i.e., monthly variation regarding the difference between model and natural system dynamics). The latter statistical formulation, including a Gaussian process model to account for the covariance structure among the stochastic/error terms (Higdon et al., 2004), may also improve the current misrepresentation of the spring plankton dynamics (underestimation of the mid–late spring phytoplankton and zooplankton levels).

In conclusion, we outlined a methodological framework that integrates environmental mechanistic (process-based) modeling with Bayesian analysis. Through an illustration focused on the reproduction of the average seasonal patterns of a mesotrophic lake (Lake Washington), we showed that this approach provides a convenient means for characterizing uncertainty in model predictions. We also discussed several of the anticipated benefits from the Bayesian calibration that are well suited for stakeholders and policy makers when making decisions for sustainable environmental management. Sound environmental management can only result from an in-depth assessment of the political/social factors, scientific considerations, and economic impacts; the proposed methodological framework can be very useful in this direction and can facilitate decisions for rational resource allocation.

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 US EPA (STAR Grant #R830883). All the material pertinent to this study is available upon request from the first author.

REFERENCES

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

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

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

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

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

Bates, R.A., Buck, R.J., Riccomagno, E., Wynn, H.P., 1996. Experimental design and observation for large systems. J. Roy. Stat. Soc.B Met. 58, 77–94.
Bayarri, M., Berger, J.O., Higdon, D., Kennedy, M., Kottas, A., Paulo, R., Sacks, J., Cafeo, J., Cavendish, J., Tu, J., 2002. A framework for the validation of computer models. In: Pace, D., Stevenson, S. (Eds.), Proceedings of the Workshop on Foundations for V&V in the 21st Century. Society for Modeling and Simulation International.
Beck, M.B., 1987. Water-quality modelling—a review of the analysis of uncertainty. Water Resour. Res. 23, 1393–1442.
Beven, K., Binley, A., 1992. The future of distributed models–model calibration and uncertainty prediction. Hydrol. Process. 6, 279–298.
Borsuk, M.E., Stow, C.A., Reckhow, K.H., 2002. Predicting the frequency of water quality standard violations: a probabilistic approach for TMDL development. Environ. Sci. Technol. 36, 2109–2115.
Brett, M.T., Mueller, S.E., Arhonditsis, G.B., 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., Gelman, A., 1998. Alternative methods for monitoring convergence of iterative simulations. J. Comput. Graph. Stat. 7, 434–455.
Chapra, S.C., Canale, R.P., 1998. Numerical Methods for Engineers, 3rd ed. McGraw-Hill, 924 pp.
Clark, J.S., Carpenter, S.R., Barber, M., Collins, S., Dobson, A., Foley, J.A., Lodge, D.M., Pascual, M., Pielke, R., Pizer, W., Pringle, C., Reid, W.V., Rose, K.A., Sala, O., Schlesinger, W.H., Wall, D.H., Wear, D., 2001. Ecological forecasts: an emerging imperative. Science 293, 657–660.
Craig, P.S., Goldstein, M., Rougier, J.C., Seheult, A.H., 2001. Bayesian forecasting for complex systems using computer simulators. J. Am. Stat. Assoc. 96, 717–729.
Currin, C., Mitchell, T., Morris, M., Ylvisaker, D., 1991. Bayesian prediction of deterministic functions with applications to the design and analysis of computer experiments. J. Am. Stat. Assoc. 86, 953–963.
Edmondson, W.T., 1994. Sixty years of Lake Washington: a curriculum vitae. Lake Reserv. Manage. 10, 75–84.
Edwards, A.M., 2001. Adding detritus to a nutrient–phytoplankton–zooplankton model: a dynamical-systems approach. J. Plankton Res. 23, 389–413.
Edwards, A.M., Yool, A., 2000. The role of higher predation in plankton population models. J. Plankton Res. 22, 1085–1112.
Gelman, A., 2005. Prior distributions for variance parameters in hierarchical models. Bayesian Anal. 1, 1–19.
Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 1995. Bayesian Data Analysis. Chapman and Hall, New York, p. 518.
Gelman, A., Meng, X.L., Stern, H., 1996. Posterior predictive assessment of model fitness via realized discrepancies. Stat. Sinica 6, 733–807.
Gilks, W.R., Richardson, S., Spiegelhalter, D.J. (Eds.), 1998. Markov Chain Monte Carlo in Practice. Chapman & Hall/CRC, 512 pp.
Goldstein, M., Rougier, J., 2004. Probabilistic formulations for transferring inferences from mathematical models to physical systems. SIAM J. Sci. Comput. 26, 467–487.
Hegstad, B.K., More, H., 2001. Uncertainty in production forecasts based on well observations, seismic data, and production history. Soc. Petrol Eng. J. 6, 409–424.
Higdon, D., Kennedy, M., Cavendish, J.C., Cafeo, J.A., Ryne, R.D., 2004. Combining field data and computer simulations for calibration and prediction. SIAM J. Sci. Comput. 26, 448–466.
Hoeting, J.A., Madigan, D., Raftery, A.E., Volinsky, C.T., 1999. Bayesian model averaging: a tutorial. Stat. Sci. 14, 382–401.
Jansen, M.J.W., Hagenaars, T.J., 2004. In: Van Boekel, M.A.J.S., Stein, A., Van Bruggen, A.H.C. (Eds.), Calibration in a Bayesian modeling framework. In Bayesian Statistics and Quality Modeling in the Agro-Food Production Chain. Kluwer, 47–55.
Jorgensen, S.E., Bendoricchio, G., 2001. Fundamentals of Ecological Modelling, 3rd ed. Elsevier Publications, 530 pp.
Kass, R.E., Raftery, A.E., 1995. Bayes factors. J. Am. Stat. Assoc. 90, 773–795.
Kennedy, C.M., O'Hagan, A., 2001. Bayesian calibration of computer models. J. R. Stat. Soc. B 63, 425–464.
Lee, H.K.H., Higdon, D.M., Bi, Z.O., Ferreira, M.A.R., West, M., 2002. Markov random field models for high-dimensional parameters in simulations of fluid flow in porous media. Technometrics 44, 230–241.
Legendre, P., Legendre, L., 1998. Numerical Ecology, 2nd ed. Elsevier Science BV, Amsterdam, 853.
Neal, R., 1998. Suppressing random walks in Markov chain Monte Carlo using ordered over-relaxation. In: Jordan, M.I. (Ed.), Learning in Graphical Models. Kluwer Academic Publishers, Dordrecht, pp. 205–230.
Omlin, M., Reichert, P., 1999. A comparison of techniques for the estimation of model prediction uncertainty. Ecol. Model. 115, 45–59.
Poole, D., Raftery, A.E., 2000. Inference for deterministic simulation models: the Bayesian melding approach. J. Am. Stat. Assoc. 95, 1244–1255.
Reckhow, K.H., 1994. Water-quality simulation modeling and uncertainty analysis for risk assessment and decision making. Ecol. Model. 72, 1–20.
Reckhow, K.H., Chapra, S.C., 1999. Modeling excessive nutrient loading in the environment. Environ. Pollut. 100, 197–207.
Reichert, P., Omlin, M., 1997. On the usefulness of overparameterized ecological models. Ecol. Model. 95, 289–299.
Reichert, P., Schervish, M., Small, M.J., 2002. An efficient sampling technique for Bayesian inference with computationally demanding models. Technometrics 44, 318–327.
Richey, J.E., 1979. Patterns of phosphorus supply and utilization in Lake Washington and Findley Lake. Limnol. Oceanogr. 24, 906–916.
Robert, C.P., Casella, G., 1999. Monte Carlo Statistical Methods. Springer-Verlag, 507 pp.
Saltelli, A., Chan, K., Scott, E.M., 2000. Sensitivity Analysis. Wiley & Sons, 475 pp.
Scheffer, M., Rinaldi, S., Kuznetsov, Y.A., van Nes, E.H., 1997. Seasonal dynamics of *Daphnia* and algae explained as a periodically forced predator–prey system. Oikos 80, 519–532.
Spiegelhalter, D., Thomas, A., Best, N., 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., Wolpert, R.L., 1996. Bayesian model for fate and transport of polychlorinated biphenyl in upper Hudson River. J. Environ. Eng. ASCE 122, 341–349.
Steinberg, L.J., Reckhow, K.H., Wolpert, R.L., 1997. Characterization of parameters in mechanistic models: a case study of a PCB fate and transport model. Ecol. Model. 97, 35–46.
Stow, C.A., Roessler, C., Borsuk, M.E., Bowen, J.D., Reckhow, K.H., 2003. Comparison of estuarine water quality models for total maximum daily load development in Neuse River Estuary. J. Water Res. Pl. ASCE 129, 307–314.
Stow, C.A., Lamon, E.C., Qian, S.S., Schrank, C.S., 2004. Will Lake Michigan lake trout meet the Great Lakes strategy 2002 PCB reduction goal? Environ. Sci. Technol. 38, 359–363.
Van Oijen, M., Rougier, J., Smith, R., 2005. Bayesian calibration of process-based forest models: bridging the gap between models and data. Tree Physiol. 25, 915–927.
Walters, C.J., 1986. Adaptive Management of Renewable Resources. McMillan, New York.
Zhang, W., Arhonditsis, G.B., submitted for publication. Predicting the frequency of water quality standard violations using Bayesian calibration of eutrophication models.