



# Bayesian calibration of mathematical models: Optimization of model structure and examination of the role of process error covariance



Maryam Ramin, George B. Arhonditsis\*

*Ecological Modelling Laboratory, Department of Physical & Environmental Sciences, University of Toronto, Toronto, Ontario M1C 1A4, Canada*

## ARTICLE INFO

### Article history:

Received 24 March 2013

Received in revised form 29 June 2013

Accepted 13 July 2013

Available online 19 July 2013

### Keywords:

Process-based modeling

Eutrophication

Bayesian inference

Process error covariance

Model averaging

Model complexity optimization

## ABSTRACT

The integration of Bayesian inference techniques with mathematical modeling offers a promising means to improve ecological forecasts and management actions over space and time, while accounting for the uncertainty underlying model predictions. In this study, we address two important questions related to the ramifications of the statistical assumptions typically made about the model structural error and the prospect of Bayesian calibration to guide the optimization of model complexity. Regarding the former issue, we examine statistical formulations that whether postulate conditional independence or explicitly accommodate the covariance among the error terms for various model endpoints. Our analysis evaluates the differences in the posterior parameter patterns and predictive performance of a limiting nutrient (phosphate)–phytoplankton–zooplankton–detritus (particulate phosphorus) model calibrated with three alternative statistical configurations. The lessons learned from this exercise are combined with those from a second comparative analysis that aims to optimize model structure. In particular, we selected three formulas of the zooplankton mortality term (linear, hyperbolic, sigmoidal) and examine their capacity to determine the posterior parameterization as well as the reproduction of the observed ecosystem patterns. Our analysis suggests that the statistical characterization of the model error as well as the mathematical representation of specific ecological processes can be influential to the inference drawn by a modeling exercise. Our findings could be useful when selecting the most suitable statistical framework for model calibration and/or making informative decisions about model structure optimization. In the absence of adequate prior knowledge, we also advocate the use of Bayesian model averaging for obtaining weighted averages of the forecasts from different model structures and/or statistical descriptions of the process error terms.

© 2013 Elsevier B.V. All rights reserved.

## 1. Introduction

The rigorous analysis of decision problems in eutrophication management requires fundamental understanding of the biogeochemical cycles; specification of objective functions for evaluating alternative management strategies; predictive models of ecosystem dynamics formulated in terms of variables relevant to management objectives; a finite set of alternative management plans, including any limitations on their use; and a monitoring program to follow system response to restoration actions (Arhonditsis et al., 2011). An inherently difficult task in practical applications of decision theory is the impartial characterization of an objective function, which specifies the value of alternative management actions and usually accounts for benefits, costs, and conditional constraints (Dorazio and Johnson, 2003). Likewise, the predictive models aim to realistically reproduce the relevant behaviors of aquatic ecosystems that are nonlinear, complex, and are characterized by spatial, temporal, and organizational heterogeneity (Arhonditsis and Brett, 2004). Perhaps even greater challenges are posed by the uncertainty in predictions of management outcomes. This uncertainty may

stem from inadequate control of management actions, incomplete knowledge of system behavior, errors in measurement and sampling of aquatic ecosystems, natural variability, and model structural or parametric uncertainty (Arhonditsis et al., 2007; Borsuk et al., 2004; Walters and Holling, 1990). Failure to recognize and account for these sources of uncertainty can severely compromise management performance, and in some cases, has led to catastrophic environmental and economic losses (Williams et al., 1996; Walters, 1986). However, the general lack of uncertainty estimates for most eutrophication models, the arbitrary selection of higher – and often unattainable – threshold values for water quality standards as a hedge against unknown forecast errors, risky model-based management decisions and unanticipated system responses are still the typical management practice (Arhonditsis, 2009; Arhonditsis et al., 2007; 2008a,b).

Given this ominous context, there has been a growing interest in the theory of stochastic decision processes and the development of practical methods that can explicitly accommodate the uncertainty in the response of environmental systems to both controlled and uncontrolled factors (Dorazio and Johnson, 2003). In this regard, particular emphasis has been placed on the implementation of Bayesian inference methods that enable the explicit consideration of model uncertainty, can be engaged with the policy practice of adaptive management, and have the

\* Corresponding author. Tel.: +1 416 208 4858; fax: +1 416 287 7279.

E-mail address: [georgea@utsc.utoronto.ca](mailto:georgea@utsc.utoronto.ca) (G.B. Arhonditsis).

ability to update and improve model predictions and management actions in space and time (Kennedy and O'Hagan, 2001; Zhang and Arhonditsis, 2008). The Bayesian inference is consistent with the scientific process of progressive learning and offers a natural mechanism for sequentially updating beliefs (specified in terms of model parameters) every time new data are collected from the system and for predicting the consequences of future management actions, while properly accounting for uncertainty in the updated beliefs (Arhonditsis et al., 2008a). Recent research has also shown that the Bayesian paradigm can effectively alleviate problems of spatiotemporal resolution mismatch among different submodels of integrated environmental modeling systems, overcome the conceptual or scale misalignment between processes of interest and supporting information, exploit disparate sources of information that differ with regard to the measurement error and resolution, and accommodate tightly intertwined environmental processes operating at different spatiotemporal scales (Boone et al., 2012; Hooten et al., 2011; Qian et al., 2010; Wikle, 2003; Wikle et al., 1998; Zhang and Arhonditsis, 2009).

Several recent studies have attempted to demonstrate the benefits of Bayesian inference techniques in the context of model-based water quality management. For example, Arhonditsis et al. (2007; 2008a,b; 2011) introduced a Bayesian calibration scheme using a wide range of complexity mathematical models and statistical formulations that explicitly accommodate measurement error, parameter uncertainty, and model structure imperfection. In particular, the statistical characterization of the calibration framework was based on one of the following assumptions: (i) a “perfect” model structure along with additive (or multiplicative) measurement error; (ii) a simulator that imperfectly represents the dynamics of the natural system and the process error is invariant with the input conditions, i.e., the difference between model and system dynamics was assumed to be constant over the annual cycle for each state variable; and (iii) a model structure that imperfectly represents the dynamics of the environmental system but the corresponding process error varies with the input conditions, i.e., time variant error terms were specified for each state variable. The former formulation postulates that the model misfit is solely caused by the error associated with the data, whereas the latter ones also consider errors in the model structure, e.g., missing key ecological processes, misspecified forcing functions, and erroneous mathematical expressions. It should also be noted that, aside from the analytical/sampling error, the term measurement error also reflects the notion that the observational data are just a “snapshot” of the real system, an instantaneous record of few components from numerous complex and interactive processes that depending on the sampling network used, the ecosystem modeled and the questions addressed, can form a pragmatic basis for evaluating model performance (Arhonditsis and Brett, 2004). The characterization of the uncertainty underlying the model parameters prior to model calibration (prior parameter distributions) was based on field observations from the studied system, laboratory studies, literature information, and expert judgment using the protocol presented by the Steinberg et al. (1997) study. The Bayesian calibration framework can then be used to quantify the information the data contain about model inputs, to offer insights into the covariance structure among parameter estimates, and to obtain predictions along with credible intervals for model outputs.

A common denominator of the aforementioned statistical formulations was the postulation of conditional independence among the error terms for various model endpoints. Striving for simplicity, this strategy offers a convenient statistical description of the “model calibration” problem, but profoundly downplays the observed covariance patterns among interconnected ecosystem variables, e.g., nutrients-phytoplankton-zooplankton. The question arising though is to what extent this pragmatic approach introduces a systematic bias in the model parameterization and may affect the capacity of the modeling exercise to support robust predictive statements. To this end, our analysis evaluates the posterior parameter patterns and predictive

performance of a limiting nutrient-phytoplankton-zooplankton-detritus model when the Bayesian calibration framework explicitly accommodates the covariance of the error terms associated with different state variables. We synthesize the lessons learned from this exercise with the findings of a second comparative analysis that aims to optimize model structure. In particular, we selected three formulas of the zooplankton mortality term (linear, hyperbolic, sigmoidal) and examine their capacity to determine the posterior parameterization as well as the reproduction of the observed patterns. Our intent is to illustrate the variety of options along with the critical decisions involved when selecting the most suitable statistical framework for model calibration and/or the optimal model structure. It is our belief that our case study will offer – much needed – prescriptive guidelines for the effective integration of Bayesian inference with process-based models.

## 2. Methods

### 2.1. Case study

The study site for our modeling work was the Hamilton Harbour, Ontario, Canada, a large embayment with long history of eutrophication problems primarily manifested as excessive algal blooms, low water transparency, predominance of toxic cyanobacteria, and low hypolimnetic oxygen concentrations during the late summer (Hiriart-Baer et al., 2009; Ramin et al., 2011). Since the mid 80s, when the Harbour was identified as one of the 43 Areas of Concern (AOC) in the Great Lakes area, the Hamilton Harbour Remedial Action Plan (RAP) was formulated through a variety of government, private sector, and community participants to provide the framework for actions aimed at restoring the Harbour environment (Hall et al., 2006). The foundation of the remedial measures and the setting of water quality goals reflect an ecosystem-type approach that considers the complex interplay between abiotic variables and biotic components pertinent to its beneficial uses (Charlton, 2001). The drastic nutrient loading reduction has historically played a central role in the restoration efforts, although the determination of the critical levels has been a thorny issue as the population growth and increasing urbanization accentuate the pressure for expansion of the local wastewater treatment plants (WWTPs) (Charlton, 2001).

Recent modeling work suggests that the water quality goals for TP levels  $<20 \mu\text{g L}^{-1}$ , chlorophyll *a* concentrations between  $5\text{--}10 \mu\text{g L}^{-1}$  and water clarity  $>3 \text{ m}$  will likely be met, if the proposed phosphorus loading reductions at the level of  $142 \text{ kg day}^{-1}$  are actually achieved (Gudimov et al., 2010; 2011; Ramin et al., 2011). However, it was emphasized that the predictive capacity of any modeling exercise in the Harbour is conditional upon the credibility of the contemporary nutrient loading estimates, which are uncertain and appear to inadequately account for the contribution of non-point sources, episodic meteorological events (e.g., spring thaw, intense summer storms), and short-term variability at the local WWTPs (Gudimov et al., 2011). A follow-up analysis considered the fact that there is no true model of an ecological system and used an averaging scheme for obtaining weighted averages of the forecasts from two models of different complexity (Ramin et al., 2012). Two important unknown factors were identified that can potentially modulate the response of the system to the exogenous nutrient loading reduction and may shape the duration of the transient phase as well as the system resilience in the “post-recovery” era. First, the dynamics of phosphorus in the sediment-water column interface are still poorly understood, and thus the historical notion that the internal loading in the Harbour is minimal may be inaccurate. Second, the lack of fundamental knowledge of the regulatory factors of herbivorous zooplankton abundance and composition, even though existing evidence suggests that a thriving zooplankton community can be instrumental for achieving faster recovery rates in the Harbour. The latter prospect highlights a central conclusion drawn from the recent modeling work

that the bottom–up (i.e., nutrient loading reduction) approach historically followed in the area was sufficient to bring the system in its present state, but any further improvements should be sought in the context of a combined bottom–up and top–down control (Gudimov et al., 2011; Ramin et al., 2011).

## 2.2. Mathematical model

We used a simple mathematical model that considers the interplay among the limiting nutrient (phosphate), phytoplankton, zooplankton, and detritus (particulate phosphorus), also known as NPZD model in the literature. The spatial segmentation of the model consists of three compartments representing the epilimnion, thermocline, and hypolimnion of the system. Detailed presentation of the mathematical equations and various assumptions can be found elsewhere (Arhonditsis et al., 2007; 2008a,b; Ramin et al., 2012), and thus we provide here a brief description of the mechanistic foundation of the model. The equation for phytoplankton biomass accounts for production, losses due to basal metabolism, herbivorous zooplankton grazing, and settling. Phytoplankton growth is directly linked to the ambient phosphorus concentrations without explicit consideration of the control exerted by the intracellular storage strategies. Phytoplankton basal metabolic losses include all internal processes that decrease algal biomass as well as natural mortality. The phosphate equation considers the phytoplankton uptake, the gains due to zooplankton excretion/predation, the bacteria-mediated mineralization of detritus, and the net diffusive fluxes between adjacent compartments. The detritus equation takes into account the contributions from phytoplankton respiration and zooplankton excretion, and the losses due to bacteria-mediated mineralization and settling. The effects of the seasonal temperature cycle on phosphate diffusion and sediment forcing are described by a trigonometric function (Arhonditsis et al., 2007). Phosphorus release from the sediments in the three spatial segments was represented by normal probability distributions, founded upon empirical estimates from previous studies in the Harbour (Azcue et al., 1998; Mayer and Manning, 1990), which were then independently updated by the Bayesian calibration exercise.

The zooplankton biomass equation considers zooplankton growth and losses due to natural mortality and predation. Zooplankton feeds upon phytoplankton and detritus with kinetics described by the Holling Type III function. Contrary to our earlier work (Law et al., 2009), the palatability of the two food sources ( $\omega$ ) is treated as a stochastic node assigned a prior distribution and subjected to updating by the calibration dataset. The zooplankton mortality/higher predation closure term can have a strong influence on the dynamics of plankton models (Edwards and Yool, 2000). In this study, we evaluated three formulas of the zooplankton mortality term and examined their capacity to influence the posterior parameterization and subsequently match the calibration data. The linear form postulates a mortality rate independent of zooplankton concentration and may be interpreted as representing a predator whose biomass does not fluctuate (Edwards and Brindley, 1999). This functional form is used in cases where there is not sufficient information to justify the use of more complex mathematical expressions. On the other hand, both the hyperbolic and sigmoidal forms are characterized by a plateau at high zooplankton concentrations, representing satiable predators, e.g., long handling times for the food items, upper bounds on predator density due to finite carrying capacity of the habitat or due to direct interference among the predator themselves (Edwards and Yool, 2000). The sigmoidal differs from the hyperbolic form in its predictions at low zooplankton concentrations, in that the former response leads to lower rates of predation than the latter one (Malchow, 1994). As zooplankton density increases though, the sigmoidal equation predicts faster increase in predation rate and the overall response resembles to a “S-shaped” pattern. This kind of predation mimics a predator with threshold behavior that switches among different prey types; namely, a predator that targets a prey only if its density exceeds a certain level to make a net biomass gain for the effort

expended (Edwards and Yool, 2000). A fraction of the zooplankton grazing is assimilated and fuels growth, another fraction is excreted as phosphate, while the remaining fraction represents the fecal pellets contributing to the detritus pool. Finally, we assumed a unimodal response of the planktonic processes on temperature seasonal variability modeled by a Gaussian-like probability curve (Arhonditsis and Brett, 2005).

## 2.3. Statistical formulations

All of the statistical formulations examined are founded upon the assumption that the plankton model is an imperfect simulator of the system dynamics and the structural error is constant over the annual cycle for each state variable. The uncertainty associated with the dataset is also accounted for with a data quality submodel (Wellen et al., 2012). This component of our framework stipulates that each observation from the system is a random draw from a normal distribution, in which the mean value represents the (latent) error-free observation (also referred to as “true value”) and the variance is associated with the sampling error or other sources of uncertainty, e.g., variability in time/space (Carroll et al., 2006). In particular, we assumed that the monthly standard deviations of the modeled water quality variables were 25% of the corresponding mean monthly values, a fraction that comprises both analytical error and the inter-annual variability in the Hamilton Harbour (Hiriart-Baer et al., 2009; Ramin et al., 2011). Under the assumption of conditional independence among the model error terms, an observation  $i$  for the state variable  $j$ ,  $y_{ij}$ , can be described as:

$$\begin{aligned} y_{ij} &\sim N(\hat{y}_{ij}, \sigma_{ijobs}) \\ \hat{y}_{ij} &\sim N(f(\theta, x_i, y_0), \sigma_j) \\ \theta &\sim MVLN(\theta_\mu, \Sigma_\theta) \quad y_0 \sim MVN(y_{0\mu}, \Sigma_0) \\ \sigma_{ijobs}^2 &= (0.25 \cdot y_{ij})^2 \cdot 1/\sigma_j^2 \sim \text{Gamma}(0.001, 0.001) \\ &i=1, 2, 3, \dots, n \text{ and } j=1, \dots, m \end{aligned} \quad (1)$$

where  $\hat{y}_{ij}$  represents the latent “true value” used to parameterize the process-based models;  $\sigma_{ijobs}$  corresponds to the observation error;  $f(\theta, x_i, y_0)$  denotes the plankton models that consider linear (Model 4), sigmoidal (Model 5), and hyperbolic (Model 6) zooplankton mortality (closure) terms;  $\sigma_j$  is the time-independent, variable-specific process (structural) error term;  $x_i$  is a vector of time-dependent control variables (e.g., boundary conditions, forcing functions) describing the prevailing nutrient loading/weather conditions; the vector  $\theta$  is a time independent set of the calibration model parameters;  $\theta_\mu$  indicates the vector of the mean values of  $\theta$  in logarithmic scale;  $\Sigma_\theta = I_l \cdot \sigma_\theta^T \cdot \sigma_\theta$  and  $\sigma_\theta = [\sigma_{\theta_1}, \dots, \sigma_{\theta_l}]^T$  corresponds to the vector of the shape parameters of the  $l$  lognormal distributions (standard deviation of  $\log \theta$ ), where  $l = 16$  and  $17$  depending on the model selected;  $y_0$  corresponds to the vector of the values of the state variables at the initial time point  $t_0$  (initial conditions); the vector  $y_{0\mu} = [y_{1,1}, \dots, y_{1,m}]^T$  corresponds to the January values of all the state variables of the model;  $\pi = 12$  (simple model: 4 state variables  $\times$  3 spatial compartments);  $m (=6)$  corresponds to the number of state variables for which empirical information from the system exists and thus contribute to the model likelihood (phosphate, phytoplankton, zooplankton, and particulate phosphorus in the epilimnion along with hypolimnetic phosphate and particulate phosphorus);  $n$  is the number of observations in time used to calibrate the model (12 average monthly values);  $MVLN$  and  $MVN$  represent the multivariate lognormal and multivariate normal distributions, respectively. The characterization of the prior density of their initial values was based on the assumption of a Gaussian distribution with mean values derived from the January monthly averages and moderately diffuse standard deviations, specified as 25% of the mean value for each state variable  $j$ ; i.e.,  $\Sigma_0 = I_\pi \cdot (0.25)^2 \cdot y_{0\mu}^T \cdot y_{0\mu}$ . Notably, when flat

priors were assigned to the initial conditions, the inference drawn remained practically unaltered, although the posterior uncertainty associated with the initial conditions was somewhat greater.

In the next step, our Bayesian calibration approach explicitly accommodated the covariance of the error terms associated with the state variables of the mathematical model with the linear zooplankton mortality. We examined three statistical formulations that assumed (i) no prior knowledge on the process error values but high confidence on our knowledge of their correlation patterns (*Model 1*), (ii) some prior knowledge on the variance and covariance of the error terms (*Model 2*), and (iii) high level of confidence on both process error variance and covariance values (*Model 3*). Based on these specifications, an observation  $y_{ij}$  can be described as follows:

$$\begin{aligned}
 y_{ij} &\sim N(\hat{y}_{ij}, \sigma_{ijobs}) \\
 \hat{y}_{ij} &\sim MVN(f(\theta, x_i, y_0), \Sigma_M) \\
 \theta &\sim MVLN(\theta_\mu, \Sigma_\theta) \quad y_0 \sim MVN(y_{0\mu}, \Sigma_0) \\
 \sigma_{ijobs}^2 &= (0.25 \cdot y_{ij})^2 \\
 \Sigma_M &= \rho \cdot \sigma^T \cdot \sigma \\
 \rho &= \begin{bmatrix} 1, r_{12}, \dots, r_{1j-1}, r_{1j} \\ \dots \\ r_{j1}, r_{j2}, \dots, r_{j-1, j-1}, 1 \end{bmatrix}, \quad \sigma = [\sigma_1, \dots, \sigma_j]^T \quad (2) \\
 1/\sigma_j^2 &\sim \text{Gamma}(0.001, 0.001) && \text{(Model 1)} \\
 \Sigma_M^{-1} &\sim \text{Wishart}(\Omega, n) && \text{(Model 2)} \\
 1/\sigma_j^2 &\sim \text{Gamma}(\alpha_1, \alpha_2) && \text{(Model 3)} \\
 i &= 1, 2, 3, \dots, n \text{ and } j = 1, \dots, m
 \end{aligned}$$

where  $\Sigma_M$  represents the process (structural) error covariance matrix, which provides the basis for the three statistical formulations: (i) the error variances are treated as stochastic nodes with no prior knowledge on the respective values, while the correlation coefficients are assumed to be known and thus remained fixed during the model updating. Specifically, the error precision terms,  $1/\sigma_j^2$  are assigned flat gamma priors and the process error correlation values are specified by the matrix  $\rho$ , reflecting the spatio-temporal correlation patterns among the corresponding water quality variables in the system, i.e., phosphate, chlorophyll *a*, zooplankton biomass, and particulate phosphorus (*Model 1*); (ii) the entire process error covariance matrix is treated stochastically. Namely, the precision (inverse covariance) matrix  $\Sigma_M^{-1}$  is assigned a Wishart prior (*Model 2*), in which the scale matrix  $\Omega$  represents an assessment of the order of magnitude of the covariance matrix among the water quality variables, as has been historically manifested in the system (Hiriart-Baer et al., 2009). To represent some confidence on the existing information, we chose eight degrees of freedom for this distribution ( $n = 8$ ), which is greater than the rank of the matrix ( $=6$ ); and (iii) informative priors for the individual error terms  $\sigma_j$  are combined with the empirical correlation matrix  $\rho$  (*Model 3*). In particular, we assumed that the average error precisions are equal to the values typically reported for the same state variables in the modeling literature in eutrophic systems (see Arhonditsis and Brett, 2004). That is, the mean square error for phosphate, particulate phosphorus, phytoplankton and zooplankton biomass were assumed to be equal to  $1.5 \mu\text{g P L}^{-1}$ ,  $2 \mu\text{g P L}^{-1}$ ,  $200 \mu\text{g C L}^{-1}$ , and  $50 \mu\text{g C L}^{-1}$ , respectively, while the corresponding uncertainty (standard deviation) was set half to their mean values.

Sequences 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 over-relaxation to control the serial correlation of the MCMC samples (Neal, 1998). As originally proposed by Arhonditsis et al. (2007), the present Bayesian parameter estimation is based on two parallel chains with starting points: (i) a vector that consists of the mean values of the prior parameter distributions, and (ii) a vector based on a preliminary deterministic calibration of

the two models. The models were run for 30,000 iterations and convergence was assessed with the modified Gelman–Rubin convergence statistic (Brooks and Gelman, 1998). The accuracy of the posterior parameter values was inspected by assuring that the Monte Carlo error for all parameters was less than 5% of the sample standard deviation. Our framework is implemented in the WinBUGS Differential Interface (WBDiff), an interface that allows numerical solution of systems of ordinary differential equations within the WinBUGS software. Aside from the differences in the central tendency and the underlying uncertainty, we also evaluated the degree of updating between parameter priors and posteriors by assessing the changes in the shape of the corresponding distributions using the delta index (Endres and Schindelin, 2003; Hong et al., 2005). The delta index measures the distance between two probability distributions:

$$\delta_{\theta_i} = \sqrt{\int \left( \pi(\theta_i) \log \frac{2\pi(\theta_i)}{\pi(\theta_i) + \pi(\theta_i|D)} + \pi(\theta_i|D) \log \frac{2\pi(\theta_i|D)}{\pi(\theta_i) + \pi(\theta_i|D)} \right) d\theta} \quad (3)$$

where  $\pi(\theta_i)$  and  $\pi(\theta_i|D)$  represent the marginal prior and posterior distributions of parameter  $\theta_i$ , respectively. This metric is equal to zero if there is no difference between the two distributions, and equal to  $\sqrt{2 \log 2}$  if there is no overlap between the two distributions. All delta index values will be presented as percentages of this maximum value.

### 3. Results–discussion

#### 3.1. How robust is the posterior parameterization among the different model configurations?

The two MCMC sequences for all the models examined converged rapidly ( $\approx 4500$ – $5000$  iterations) and the statistics reported were based on the last 25,000 draws by keeping every 20th iteration (thin = 20). Posterior mean values and standard deviations of the calibration parameter vector of the statistical formulation that treats all the entries of the error covariance matrix as stochastic nodes (*Model 2*) and the model that considers sigmoid zooplankton mortality (*Model 5*) are provided in Tables 1 and Figs. 1–2. Similarly, the central tendencies of the rest of the model configurations along with the underlying uncertainty are presented in Table 1 and Figs. 1–4 in the electronic Supplementary material. Relative to the prior parameter specification, the posterior statistics generally indicate that a substantial amount of knowledge was gained for most of the parameters after the model updating. Characteristic examples were the distinct shifts of the most likely values of the zooplankton half saturation constant ( $\mu$ ), respiration rate ( $r$ ), detritus sinking rate ( $\psi$ ), detritus mineralization rate ( $\varphi$ ), and the zooplankton preference for detritus relative to phytoplankton ( $\omega$ ). Likewise, several of the marginal posteriors were characterized by significantly narrower standard deviations compared to those specified prior to model calibration, such as the mineralization rate ( $\varphi$ ), detritus sinking rate ( $\psi$ ), zooplankton preference for detritus relative to phytoplankton ( $\omega$ ), zooplankton grazing rate ( $\lambda$ ), and phosphorus half saturation constant ( $K_p$ ). Simply, the consideration of the calibration data suggests that the kinetics of zooplankton grazing are somewhat slower than what was assumed based on prior literature information, zooplankton more heavily relies upon the algal standing stock in the system (i.e., higher selectivity for algal food items), the mineralization rates of particulate phosphorus are relatively slower than what is typically experienced in eutrophic systems, the particle settling rate varies between  $0.4$ – $0.8 \text{ m day}^{-1}$ , and the phytoplankton characterization tends towards higher respiration rates and somewhat faster phosphorus kinetics. On the other hand, there were several parameters with posterior distributions that remained – almost consistently – unaltered, suggesting that the dataset used did not offer substantial insights into the parameterization of the phytoplankton settling rate ( $s$ ), the



**Table 1**

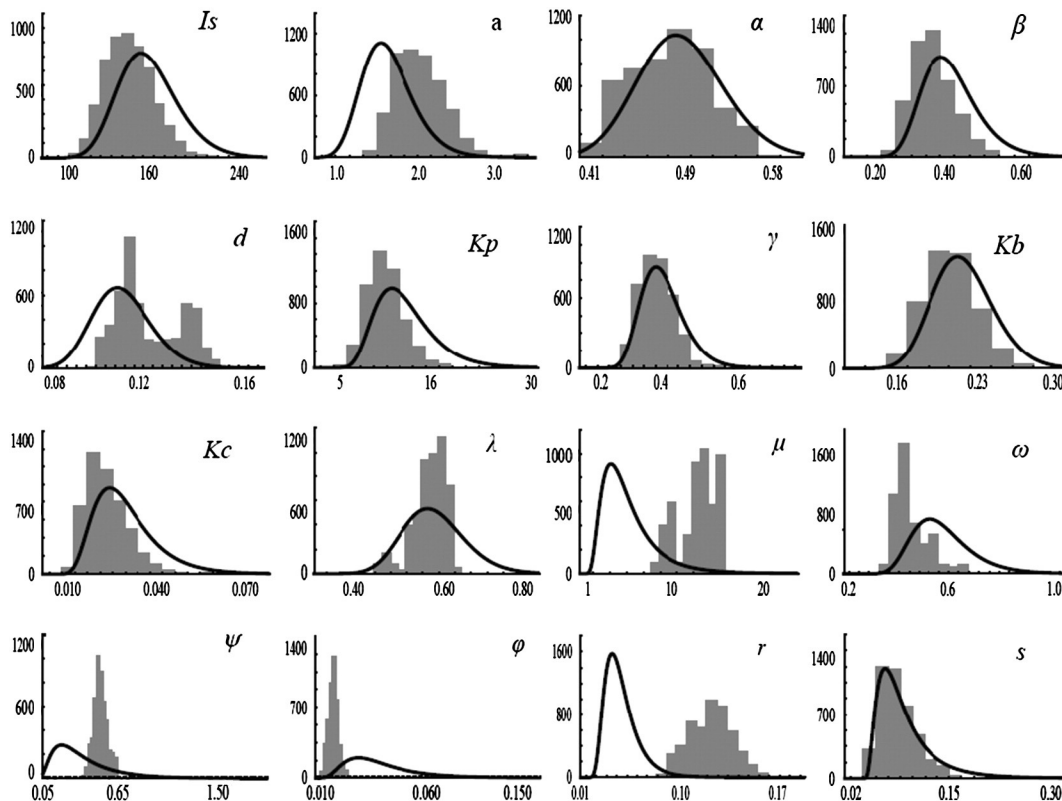
Parameter definitions and posterior estimates of the mean values and standard deviations of the stochastic nodes of statistical formulations that assign a Wishart prior to the error covariance matrix (*Model 2*), and a model that considers sigmoid (*Model 5*) zooplankton mortality.

Parameters	Description	Units	Priors		Posteriors			
			Min	Max	<i>Model 2</i>		<i>Model 5</i>	
					Mean	SD	Mean	SD
<i>a</i>	Maximum phytoplankton growth rate	day <sup>-1</sup>	1.000	3.000	2.221	0.364	2.243	0.368
<i>d</i>	Zooplankton mortality rate	day <sup>-1</sup>	0.080	0.160	0.128	0.016	0.115	0.014
<i>Kp</i>	Half-saturation constant for PO <sub>4</sub> uptake	mg P m <sup>-3</sup>	5.000	30.00	9.988	2.854	9.994	2.537
<i>r</i>	Phytoplankton respiration rate	day <sup>-1</sup>	0.010	0.100	0.119	0.026	0.099	0.026
<i>s</i>	Phytoplankton sinking loss rate	m day <sup>-1</sup>	0.010	0.300	0.064	0.045	0.082	0.058
<i>μ</i>	Zooplankton grazing half-saturation coefficient	mg P m <sup>-3</sup>	1.000	20.00	13.54	2.285	15.24	4.016
<i>φ</i>	Detritus mineralization rate	day <sup>-1</sup>	0.010	0.150	0.013	0.005	0.012	0.004
<i>ψ</i>	Detritus sinking rate	m day <sup>-1</sup>	0.050	1.500	0.447	0.062	0.422	0.031
<i>λ</i>	Maximum zooplankton grazing rate	day <sup>-1</sup>	0.400	0.800	0.554	0.045	0.487	0.072
<i>kb</i>	Background light extinction coefficient	m <sup>-1</sup>	0.150	0.300	0.202	0.025	0.205	0.026
<i>kc</i>	Light extinction coefficient due to chlorophyll a	m <sup>2</sup> mg <sup>-1</sup>	0.010	0.080	0.023	0.009	0.021	0.006
<i>α</i>	Zooplankton assimilation efficiency	–	0.400	0.600	0.479	0.036	0.461	0.031
<i>β</i>	Zooplankton excretion fraction to phosphate	–	0.200	0.700	0.319	0.070	0.337	0.082
<i>γ</i>	Zooplankton predation excretion fraction to phosphate	–	0.200	0.700	0.353	0.073	0.407	0.102
<i>ω</i>	Relative zooplankton preference for detritus compared to phytoplankton	–	0.200	1.000	0.329	0.108	0.271	0.095
<i>I<sub>s</sub></i>	Half saturation light intensity	MJ m <sup>-2</sup> day <sup>-1</sup>	100.0	250.0	142.6	22.56	146.9	22.79
<i>pred</i>	Half saturation constant for predation	mg C m <sup>-3</sup>	35	80	–	–	40.36	10.06

background light attenuation (*Kb*), the light half saturation constant (*I<sub>k</sub>*), and the zooplankton predation excretion fraction to phosphate (*γ*).

A notable feature of the posterior patterns of the three statistical formulations that explicitly accommodate the process error covariance was the bi- or multimodality trends characterizing some of the parameters, such as the zooplankton mortality rate (*d*), the maximum zooplankton grazing rate (*λ*), and the detritus sinking rate (*ψ*). Nonetheless, the correlation structure of the model parameters, as depicted by the MCMC

posterior samples revealed distinct differences among the three formulations (Tables 2–5 in electronic Supplementary material). In particular, the formulation that treats all the entries of the error covariance matrix as stochastic nodes was characterized by strong correlations among the model parameters, which in turn may reflect an incomplete coverage of their joint posterior distribution (Qian et al., 2003). However, the assessment of the changes in the shape of the parameter marginal distributions after model updating does not provide strong support of the



**Fig. 1.** Comparison between prior and posterior parameter distributions of the calibration vector of the statistical formulations that assign a Wishart prior to the process error covariance matrix (*Model 2*).

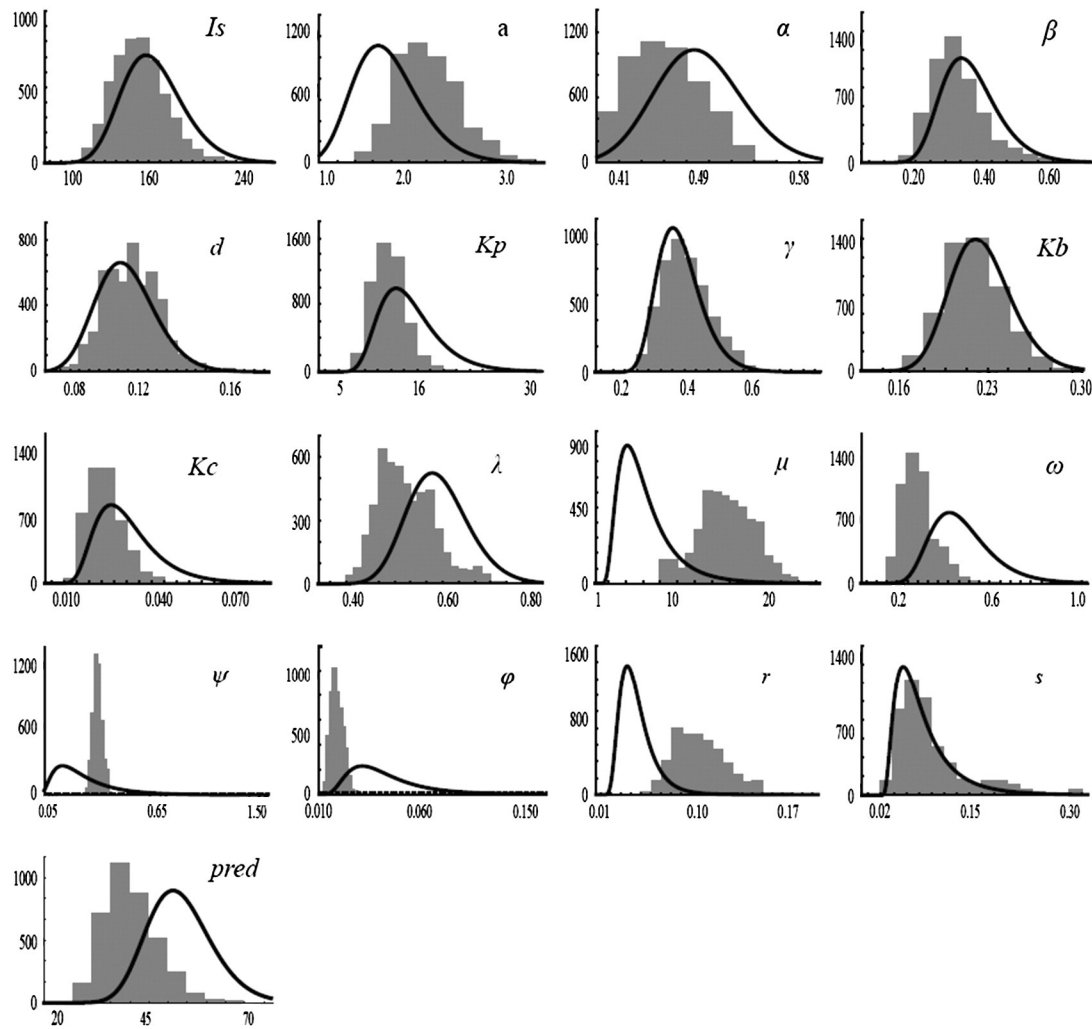


Fig. 2. Comparison between prior and posterior parameter distributions of the calibration vector of the model with the sigmoid closure term (*Model 5*).

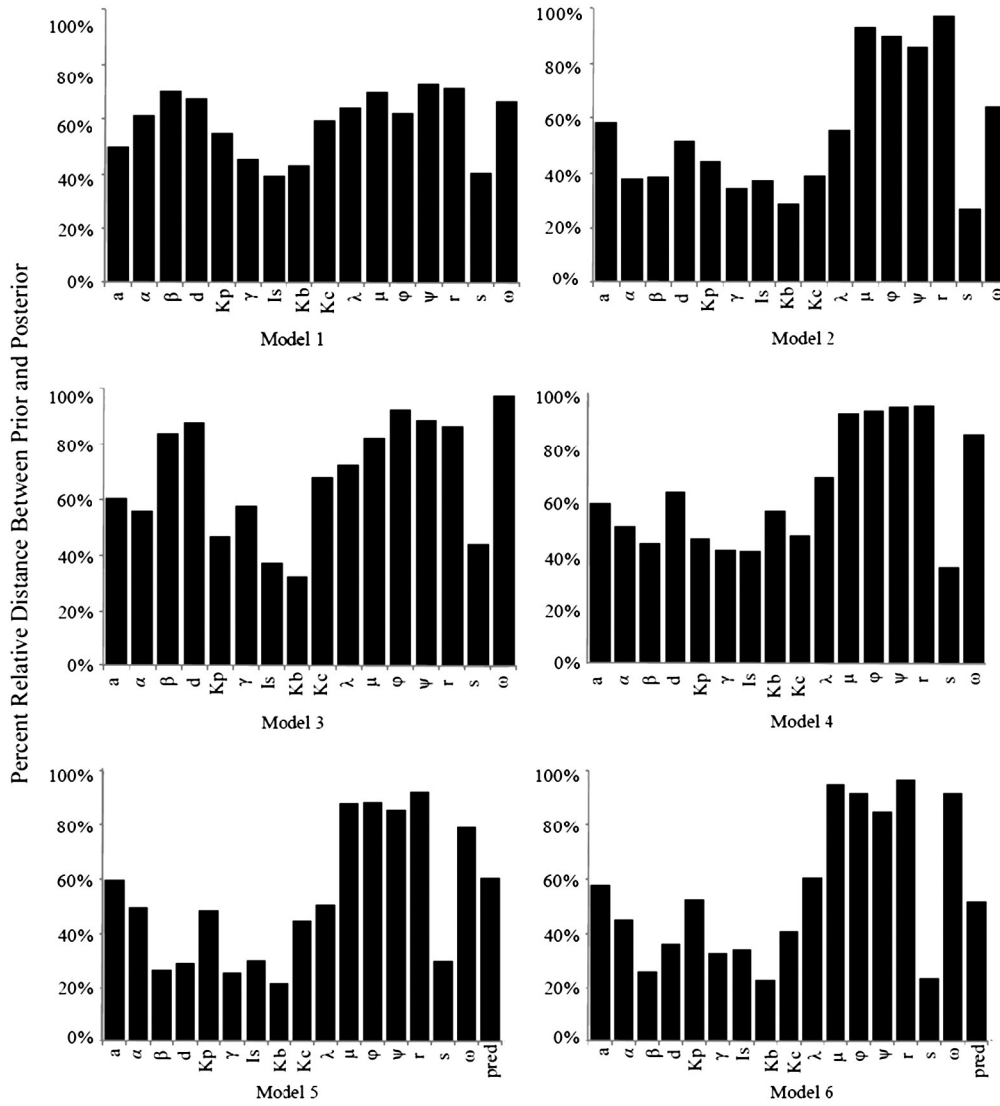
latter assertion, as the delta index values associated with *Model 2* were not distinctly different from the rest of the configurations examined (Fig. 5). Interestingly, the correlation parameter patterns were less pronounced with the formulation that assumes conditional independence among the process error terms (*Model 4*), and even weaker correlations were manifested when all the off-diagonal entries of the covariance matrix were assigned fixed correlation values of the process error terms, i.e., *Models 1* and *3* (See electronic Supplementary material).

The delta index reiterates our previous assertion that there are parameters, such as the zooplankton grazing half-saturation constant ( $\mu$ ), mineralization rate ( $\varphi$ ), detritus sinking rate ( $\psi$ ), phytoplankton respiration rate ( $r$ ), and zooplankton preference for detritus relative to phytoplankton ( $\omega$ ) that consistently demonstrate the highest distance between priors and posteriors, while other parameters, such as the background light extinction coefficient ( $kb$ ), the phytoplankton settling rate ( $s$ ), zooplankton excretion fraction to phosphate ( $\beta$ ), and maximum zooplankton grazing rate ( $\lambda$ ) are characterized by the lowest delta values across all the model configurations examined (Fig. 5). We also note the relatively uniform patterns of the delta index across the calibration parameter vector, when assigning flat gamma priors coupled with fixed correlation to the entries of the process error covariance matrix (*Model 1*). Generally, even though there were discrepancies in the marginal posterior distributions of the calibration parameters, our analysis provides evidence that neither the statistical characterization of the model error nor the closure term alone can dramatically alter the parameterization of plankton models when fitting a specific dataset.

### 3.2. What is the optimal model configuration to fit the calibration dataset?

As noted by Ramin et al. (2012), the present model structure cannot adequately reproduce the observed plankton patterns in the Hamilton Harbour, and this result was evident across all the model configurations examined. Generally, the posterior medians along with the 95% credible intervals derived from the calibration of the six model variations were close to the observed data for epilimnetic phosphate and total phosphorus, but there were substantial discrepancies with respect to the measured chlorophyll *a* concentrations, total zooplankton biomass, and hypolimnetic phosphate (Figs. 5–6 in the electronic Supplementary material). In particular, the model accurately predicts the epilimnetic phosphate and total phosphorus levels including the winter maxima and summer minima. The model closely represents the summer chlorophyll *a* concentrations ( $\approx 15\text{--}17 \mu\text{g chla L}^{-1}$ ), but the algal abundance in the fall is distinctly over-predicted ( $> 10 \mu\text{g chla L}^{-1}$ ). The two major peaks of the zooplankton biomass are underestimated, and the model does not reproduce the phosphate hypolimnetic accumulation in the summer.

Despite the apparent similarities in their performance though, the posterior estimates of the process error terms paint a somewhat different picture in that there were some distinct differences among the different model configurations (Table 2). First, the statistical formulation that treats all the entries of the error covariance matrix as stochastic nodes (*Model 2*) appears to be characterized by significantly lower error estimates for several state variables, such as the phytoplankton abundance, zooplankton biomass, and epilimnetic particulate phosphorus.



**Fig. 3.** Assessment of the changes in the shape of parameter distributions of the six model configurations examined. The delta index is equal to zero if there is no difference between the two distributions, and equal to  $\sqrt{2\log 2}$  if there is no overlap between the two distributions. All delta values are presented as percentages of this maximum value.

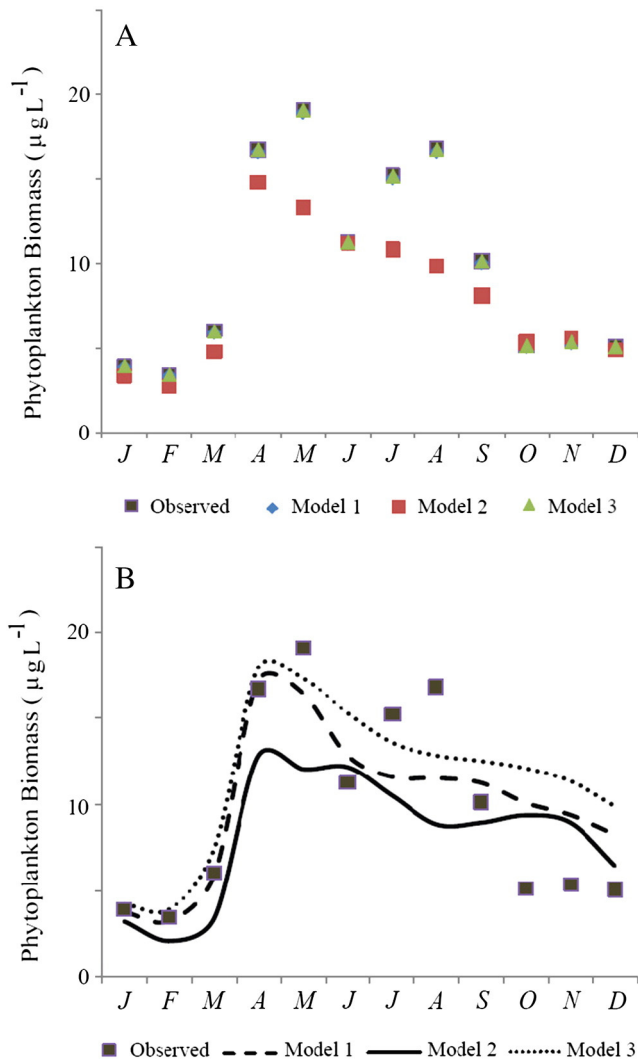
However, this finding should be interpreted with caution as the incorporation of the measurement/observation error submodel alters the conventional meaning of the structural error. Specifically, our calibration framework considers both the discrepancies between measured and “true” data as well as between “true” values and model outputs (Carroll et al., 2006). As such, the likelihood of the observation  $i$  for the state variable  $j$  given our mechanistic model is the product of the likelihood of the two levels of our hierarchical configuration:

$$\begin{aligned}
 & p(y_{ij}|\hat{y}_{ij}) \times p(\hat{y}_{ij}|f(\theta, x_i, y_0)) = \\
 & \frac{1}{\sqrt{2\pi}\sigma_{ijobs}} \exp\left(-\frac{(y_{ij}-\hat{y}_{ij})^2}{2\sigma_{ijobs}^2}\right) \times \\
 & \frac{1}{\sqrt{2\pi}\sigma_{j|-j}} \exp\left(-\frac{(\hat{y}_{ij}-f(\theta, x_i, y_0))^2}{2\sigma_{j|-j}^2}\right) \quad (4)
 \end{aligned}$$

$\sigma_{j|-j}^2 = \sigma_j^2 - \Sigma_{j,-j}\Sigma_{-j}^{-1}\Sigma_{-j,j}$

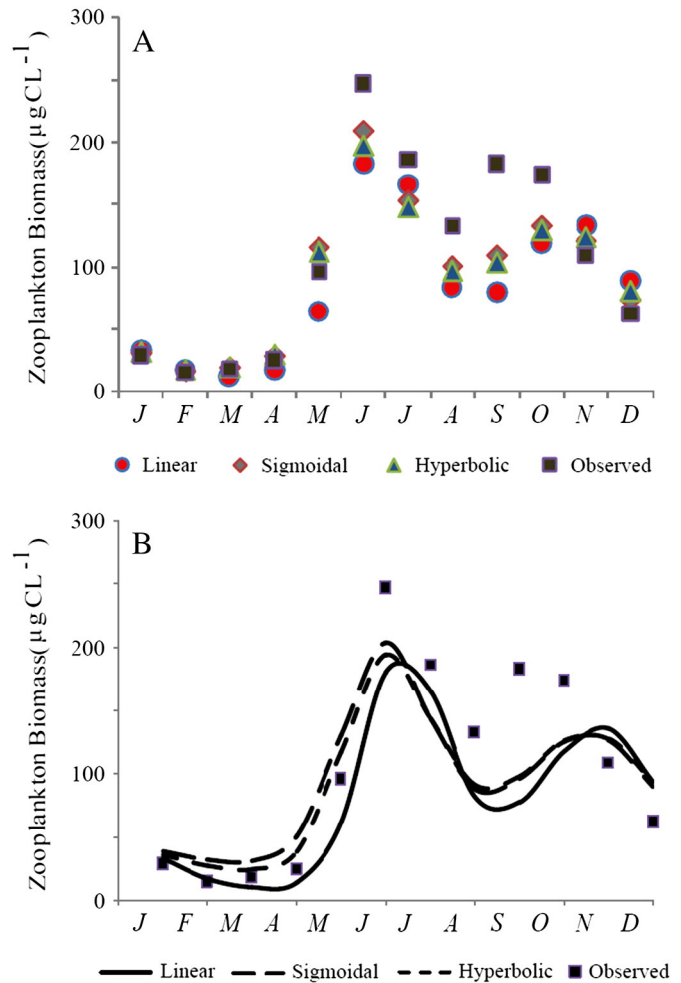
where  $\sigma_{j|-j}^2$  corresponds to the error variance for the state variable  $j$  conditional on the rest model outputs  $-j$ ; and the elements of  $\Sigma_{-j}$  and  $\Sigma_{-j,j}$  represent the process error variance for the state variables

$-j$  and the covariance between  $j$  and  $-j$ . Thus, the realistic discrimination between measurement and process errors comes with a price as it adds an additional number of  $m \times n$  stochastic nodes, thereby substantially increasing the complexity of the calibration exercise. Importantly, our analysis provides evidence that the estimation of the “true” data can differ significantly depending on the statistical description of the model error. In particular, the most complex statistical formulation (*Model 2*) results in estimates of the “true” data that significantly distance themselves from the observed values (Fig. 4). It is also important to note that the posterior predictions of the same configuration demonstrate considerable discrepancy from the measured data, and thus the aforementioned low structural error values stem from the shifts in the “true” data values rather than from a thorough exploration of the parameter space and subsequent improvement of the model fit. By contrast, the two formulations that postulate a more rigid error covariance structure (*Models 1 & 3*) are practically characterized by an overlap between “true” and observed data as well as an improved agreement with the actual observations (Fig. 4). Further, the higher delta index values for the majority of the parameters of the latter two configurations are indicative of a more efficient use of the mechanistic foundation of the modeling framework to reproduce the observed patterns (Fig. 3).



**Fig. 4.** Comparison between the observed phytoplankton biomass (chlorophyll *a*) values (black dots) and (A) the median values of the stochastic variable “true observations” or (B) the median predicted phytoplankton biomass values based on the models that assign to the process (structural) error terms flat gamma priors and prespecified correlation (*Model 1*), a Wishart prior (*Model 2*), and informative gamma priors coupled with fixed correlation (*Model 3*).

According to our analysis, the assumption of conditional independence of the process error terms does not prohibit major excursions of the stochastic node “true” values from the observed data, although the posterior error estimates are not as low as with the full stochastic treatment of the error covariance matrix (Table 2 & Fig. 5). In particular, the consideration of sigmoid zooplankton mortality appears to more effectively balance between a process-based fit of the zooplankton dynamics and a relaxation of our confidence on the credibility of the calibration dataset used (Fig. 5). It is also interesting to note that the use of linear zooplankton mortality (*Model 4*) is characterized by the lowest process error values among the closure terms examined, suggesting a greater susceptibility to the previously reported calibration pattern (Fig. 7 in electronic Supplementary material). From a management point of view, the eutrophication risk assessment statements supported by the different model configurations are fairly similar, although a careful inspection of the marginal and joint predictive posteriors of chlorophyll *a* and total phosphorus provided by the third statistical formulation (*Model 3*) are characterized by somewhat fatter tails relative to the second model (Fig. 8 in electronic Supplementary material). Importantly, the difference in the projected exceedance frequencies of the threshold value of  $40 \mu\text{g TP L}^{-1}$  (<1% versus 21%) suggests that



**Fig. 5.** Comparison between the observed zooplankton biomass data (black dots) and (A) the median values of the stochastic variable “true observations” or (B) the median predicted zooplankton biomass values based on the models with linear (circle), sigmoid (lozenge), and hyperbolic (triangle) closure terms.

the statistical characterization of the model error can alter the inference drawn regarding the assessment of prevailing environmental conditions (e.g., delisting decisions).

#### 4. Conclusions–future perspectives

In this study, we attempted to elucidate the broader implications of the statistical assumptions typically made about the model structural error as well as to evaluate the prospect of Bayesian calibration to guide the optimization of model structure. We examined statistical formulations that either postulate conditional independence or explicitly accommodate the covariance among the process error terms for various model endpoints. We also selected three mathematical expressions of the zooplankton mortality term (linear, hyperbolic, sigmoidal) and assessed their capacity to determine the posterior model parameterization and the reproduction of the observed ecosystem patterns. The basic lessons learned from the present analysis are as follows:

- Our analysis provides evidence that neither the statistical characterization of the model error nor the zooplankton mortality/higher predation term alone can significantly change the parameterization of plankton models when fitting a specific dataset. Although there were differences in the marginal posterior distributions, the ecosystem functioning postulated after the updating of the six model configurations with the training dataset was practically similar. Several parameters related to zooplankton feeding patterns, such as the zooplankton grazing half-



**Table 2**

Monte Carlo posterior estimates of the mean values and standard deviations of the process (structure) error terms of the six model configurations examined.

Parameter	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
$\sigma_{PO4epi}$	0.508	0.521	0.303	0.084	1.339	0.248	0.891	0.477	1.117	0.525	1.112	0.548
$\sigma_{PHYTepi}$	245.5	75.19	19.22	5.463	235.6	41.60	106.4	67.53	171.3	75.76	151.3	70.53
$\sigma_{ZOOpepi}$	77.35	27.16	8.659	2.374	73.59	22.10	7.634	11.53	21.57	11.86	12.58	9.982
$\sigma_{DETepi}$	9.069	9.266	0.284	0.081	1.851	0.405	0.636	0.861	0.588	0.831	0.685	0.894
$\sigma_{PO4hypo}$	1.076	0.423	0.089	0.290	0.089	0.343	1.846	0.640	2.021	0.674	2.003	0.664
$\sigma_{DEThypo}$	6.978	4.118	0.104	0.360	0.104	0.715	0.743	0.999	0.673	0.910	0.724	0.991

saturation constant ( $\mu$ ), and zooplankton preference for detritus relative to phytoplankton ( $\omega$ ), or the epilimnetic dynamics of particulate matter, such as mineralization rate ( $\varphi$ ), detritus sinking rate ( $\psi$ ), and phytoplankton respiration rate ( $r$ ), demonstrated the highest degree of change relative to our prior knowledge on their values.

- Depending on the assumptions made about the process error, the correlation patterns of the posterior parameter space were characterized by distinct differences, which in turn reflect different levels of coverage/exploration of their joint posterior distribution. Strong correlations among the model parameters were found with the statistical formulation that considers all the entries of the error covariance matrix as stochastic nodes, whereas the formulation that assumes conditional independence among the process error terms demonstrated weaker correlation patterns. The weakest parameter correlations though were found when the correlation values associated with the off-diagonal elements of the process error covariance matrix were specified prior to the model calibration.
- Our analysis revealed an interesting interplay among the error covariance structure, the zooplankton closure term, and the statistical formulation that accommodates the error associated with the calibration dataset. We provided evidence that the statistical assumption of conditional independence of the process error terms as well as the complete probabilistic treatment of the error covariance can place more emphasis on the characterization of the observation error, and thus may inefficiently capitalize on the mechanistic basis of the modeling exercise. Our study showed that these statistical configurations can overinflate the discrepancy between the latent variable “true values” and the actual data, a pattern that was not manifested when we postulated a rigid process error correlation structure. Thus, the reproduction of the observed patterns with the latter strategy appears to be based on a more meticulous scrutiny of the potential model realizations, although it is admittedly founded upon an empirical but somewhat subjective (or at least site-specific) specification of the co-dependence of the process error terms.
- Seeking for the optimal likelihood expression to quantify model structural error, our earlier work showed that the assumption of a “perfect” model structure, postulating that the misfit is solely caused by the measurement error associated with the data, usually results in narrow-shaped parameter distributions due to an over-estimation of the information content of the observations. This over-conditioning of the parameter estimates on the calibration dataset can limit the potential applicability of the model in the extrapolation domain (Arhonditsis et al., 2008a). On the other hand, the development of statistical formulations explicitly recognizing the lack of perfect simulators of natural system dynamics is a promising prospect for the Bayesian calibration framework, but a set of prescriptive guidelines is necessary to make informative decisions with respect to the selection of spatiotemporally variant or invariant process error and the explicit consideration of the covariance structure or conditional independence of the error terms. In this regard, our work suggests that the error formulation alone cannot induce dramatic changes in the model parameterization, but the differences in the inference drawn can be non-trivial depending on the dataset used, the model structure considered or the management questions addressed (Arhonditsis et al., 2007; 2008a,b). In the absence of adequate prior knowledge, an appealing alternative

could also be the use of Bayesian model averaging for obtaining weighted averages of the forecasts from different model structures and/or statistical descriptions of the process error terms (Ramin et al., 2012). In the context of ecological process-based modeling, this approach should not be viewed solely as a framework to improve our predictive tools, but rather as an opportunity to compare alternative ecological dynamics to challenge existing empirical information, to integrate across conflicting paradigms, and to compare different statistical assumptions regarding the model structural uncertainty.

- In this study, our approach is on par with the recent trend in the modeling practice to attain calibration techniques that are statistically-based and thus allow deriving probabilistic model predictions (Reichert and Schuwirth, 2012). Our findings should assist with existing efforts to combine statistical descriptions of bias in model outputs (Bayarri et al., 2007; Higdon et al., 2004; Kennedy and O'Hagan, 2001) with multiobjective optimization techniques (Efstratiadis and Koutsoyiannis, 2010; Gupta et al., 2003; Madsen et al., 2002; Reichert and Schuwirth, 2012; Vrugt et al., 2003). Future work should focus on the development of procedures to support ecological forecasts in the extrapolation model domain while impartially accounting for the process error and to efficiently accommodate the bias associated with variables for which no observations exist. While we believe that the latter issue should primarily be a factor that shapes model complexity, our experience has been that the increase of complexity remains a popular practice to reduce model bias and improve the fidelity of our simulators.

## Acknowledgments

This project has received funding support from the Natural Sciences and Engineering Research Council of Canada through a Doctoral Scholarship awarded to Maryam Ramin and a Discovery Grant awarded to George Arhonditsis. All the material pertinent to this analysis is available upon request from the corresponding author.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.ecoconf.2013.07.001>.

## References

- Arhonditsis, G.B., 2009. Useless Arithmetic? Lessons learnt from aquatic biogeochemical modelling. In: Hanrahan, G. (Ed.), *Modelling of Pollutants in Complex Environmental Systems*. ILM Publications, pp. 3–26.
- Arhonditsis, G.B., Brett, M.T., 2004. Evaluation of the current state of mechanistic aquatic biogeochemical modelling. *Marine Ecology Progress Series* 271, 13–26.
- Arhonditsis, G.B., Brett, M.T., 2005. Eutrophication model for Lake Washington (USA) Part I. Model description and sensitivity analysis. *Ecological Modelling* 187, 140–178.
- Arhonditsis, G.B., Qian, S.S., Stow, C.A., Lamon, E.C., Reckhow, K.H., 2007. Eutrophication risk assessment using Bayesian calibration of process-based models: application to a mesotrophic lake. *Ecological Modelling* 208, 215–229.
- Arhonditsis, G.B., Papanitou, D., Zhang, W.T., Perhar, G., Massos, E., Shi, M.L., 2008a. Bayesian calibration of mechanistic aquatic biogeochemical models and benefits for environmental management. *Journal of Marine Systems* 73, 8–30.
- Arhonditsis, G.B., Perhar, G., Zhang, W.T., Massos, E., Shi, M., Das, A., 2008b. Addressing equifinality and uncertainty in eutrophication models. *Water Resources Research* 44, W01420.

- Arhonditsis, G., Stremilov, S., Gudimov, A., Ramin, M., Zhang, W., 2011. Integration of Bayesian inference techniques with mathematical modelling. In: Wolanski, E., McLusky, D.S. (Eds.), *Treatise on Estuarine and Coastal Science*, 9. Academic Press, Waltham, pp. 173–192.
- Azcue, J.M., Zeman, A.J., Mudroch, A., Rosa, F., Patterson, T., 1998. Assessment of sediment Harbour Canada. *Water Science and Technology* 37, 323–329.
- Bayarri, M.J., Berger, J.O., Paulo, R., Sacks, J., Cafeo, J.A., Cavendish, J., Lin, C.H., Tu, J., 2007. A framework for validation of computer models. *Technometrics* 49, 138–154.
- Boone, E.L., Stewart-Koster, B., Kennard, M.J., 2012. A hierarchical zero-inflated Poisson regression model for stream fish distribution and abundance. *Environmetrics* 23, 207–218.
- Borsuk, M.E., Stow, C.A., Reckhow, K.H., 2004. A Bayesian network of eutrophication models for synthesis, prediction, and uncertainty analysis. *Ecological Modelling* 173, 173–188.
- Brooks, S.P., Gelman, A., 1998. Alternative methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics* 7, 434–455.
- Carroll, R.J., Ruppert, D., Stefanski, L.A., Crainiceanu, C., 2006. *Measurement Errors in Non-Linear Models: A Modern Perspective*, 2nd ed., vol. 105. Chapman & Hall/CRC.
- Charlton, M.N., 2001. The Hamilton Harbour remedial action plan: eutrophication. *Verhandlungen der Internationalischen Vereinigung Limnologie* 27, 4069–4072.
- Dorazio, R.M., Johnson, F.A., 2003. Bayesian inference and decision theory – a framework for decision making in natural resource management. *Ecological Applications* 13, 556–563.
- Edwards, A.M., Brindley, J., 1999. Zooplankton mortality and the dynamical behaviour of plankton population models. *Bulletin of Mathematical Biology* 61, 303–339.
- Edwards, A.M., Yool, A., 2000. The role of higher predation in plankton population models. *Journal of Plankton Research* 22, 1085–1112.
- Efstratiadis, A., Koutsoyiannis, D., 2010. One decade of multi-objective calibration approaches in hydrological modelling: a review. *Hydrological Sciences Journal* 55 (1), 58–78.
- Endres, D.M., Schindelin, J.E., 2003. A new metric for probability distributions. *IEEE Transactions on Information Theory* 49 (7), 1858–1860.
- Gilks, W.R., Richardson, S., Spiegelhalter, D.J., 1998. *Markov Chain Monte Carlo in Practice*. Chapman & Hall/CRC, p. 512.
- Gudimov, A., Stremilov, S., Ramin, M., Arhonditsis, G.B., 2010. Eutrophication risk assessment in Hamilton Harbour: system analysis and evaluation of nutrient loading scenarios. *Journal of Great Lakes Research* 36, 520–539.
- Gudimov, A., Ramin, M., Labencki, T., Wellen, C., Shelar, M., Shimoda, Y., Boyd, D., Arhonditsis, G.B., 2011. Predicting the response of Hamilton Harbour to the nutrient loading reductions, a modelling analysis of the “ecological unknowns”. *Journal of Great Lakes Research* 37, 494–506.
- Gupta, H.V., Bastidas, L.A., Vrugt, J.A., Sorooshian, S., 2003. In: Duan, Q., et al. (Ed.), *Multiple criteria global optimization for watershed model calibration*, in *Calibration of Watershed Models*. *Water Sci. Appl.*, vol. 6. AGU, Washington, D.C, pp. 125–132.
- Hall, J.D., O'Connor, K., Ranieri, J., 2006. Progress toward delisting a Great Lakes Area of concern: the role of integrated research and monitoring in the Hamilton Harbour Remedial Action Plan. *Environmental Monitoring and Assessment* 113, 227–243.
- 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 Journal on Scientific Computing* 26 (2), 448–466.
- Hiriart-Baer, V.P., Milne, J., Charlton, M.N., 2009. Water quality trends in Hamilton Harbour: two decades of change in nutrients and chlorophyll a. *Journal of Great Lakes Research* 35, 293–301.
- Hong, B.G., Strawderman, R.L., Swaney, D.P., Weinstein, D.A., 2005. Bayesian estimation of input parameters of a nitrogen cycle model applied to a forested reference watershed, Hubbard Brook Watershed Six. *Water Resources Research* 41, W03007.
- Hooten, M.B., Leeds, W.B., Fiechter, J., Wikle, C.K., 2011. Assessing first-order emulator inference for physical parameters in nonlinear mechanistic models. *Journal of Agricultural, Biological, and Environmental Statistics* 16, 475–494.
- Kennedy, M.C., O'Hagan, A., 2001. Bayesian calibration of computer models. *Journal of the Royal Statistical Society, B (Statistical Methodology)* 63 (3), 425–464.
- Law, T., Zhang, W., Zhao, J., Arhonditsis, G.B., 2009. Structural changes in lake functioning induced from nutrient loading and climate variability. *Ecological Modelling* 220, 979–997.
- Madsen, H., Wilson, G., Ammentorp, H.C., 2002. Comparison of different automated strategies for calibration of rainfall-runoff models. *Journal of Hydrology* 261, 48–59.
- Malchow, H., 1994. Non-equilibrium structures in plankton dynamics. *Ecological Modelling* 75, 123–134.
- Mayer, T., Manning, P.G., 1990. Inorganic contaminants in suspended solids from Hamilton Harbour. *Journal of Great Lakes Research* 16, 299–318.
- 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.
- Qian, S.S., Stow, C.A., Borsuk, M.E., 2003. On Monte Carlo methods for Bayesian inference. *Ecological Modelling* 159, 269–277.
- Qian, S.S., Cuneo, T.F., Alameddine, I., McMahon, G., Reckhow, K.H., 2010. On the application of multilevel modeling in environmental and ecological studies. *Ecology* 91, 355–361.
- Ramin, M., Stremilov, S., Labencki, T., Gudimov, A., Boyd, D., Arhonditsis, G.B., 2011. Integration of mathematical modelling and Bayesian inference for setting water quality criteria in Hamilton Harbour, Ontario Canada. *Environmental Modelling and Software* 26, 337–353.
- Ramin, M., Labencki, T., Trolle, D., Boyd, D., Arhonditsis, G.B., 2012. A Bayesian synthesis of predictions from different models for setting water quality criteria. *Ecological Modelling* 242, 127–145.
- Reichert, P., Schuwirth, N., 2012. Linking statistical description of bias to multi-objective model calibration. *Water Resources Research* 48 W09543.
- 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. *Ecological Modelling* 97, 35–46.
- Vrugt, J.A., Gupta, H.V., Bouten, W., Sorooshian, S., 2003. A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. *Water Resources Research* 39 (8), 1201.
- Walters, C.J., 1986. *Adaptive management of renewable resources*. Macmillan, New York, (374 pp.).
- Walters, C.J., Holling, C.S., 1990. Large-scale management experiments and learning by doing. *Ecology* 71 (6), 2060–2068.
- Wellen, C., Arhonditsis, G.B., Labencki, T., Dunca, B., 2012. A Bayesian methodological framework for accommodating interannual variability of nutrient loading with the SPARROW model. *Water Resources Research* 48, W10505–W.
- Wikle, C.K., 2003. Hierarchical models in environmental science. *International Statistical Review* 71 (2), 181–199.
- Wikle, C.K., Berliner, L.M., Cressie, N., 1998. Hierarchical Bayesian space-time models. *Journal of Environmental Economics and Statistics* 5 (2), 117–154.
- Williams, B.K., Johnson, F.A., Wilkins, K., 1996. Uncertainty and the adaptive management of waterfowl harvests. *Journal of Wildlife Management* 60, 223–232.
- Zhang, W., Arhonditsis, G.B., 2008. Predicting the frequency of water quality standard violations using Bayesian calibration of eutrophication models. *Journal of Great Lakes Research* 34, 698–720.
- Zhang, W., Arhonditsis, G.B., 2009. A Bayesian hierarchical framework for calibrating aquatic biogeochemical models. *Ecological Modelling* 220, 2142–2161.

**BAYESIAN CALIBRATION OF MATHEMATICAL MODELS:  
OPTIMIZATION OF MODEL STRUCTURE AND EXAMINATION OF THE ROLE  
OF PROCESS ERROR COVARIANCE**

**(Electronic Supplementary Material)**

**Maryam Ramin<sup>1</sup> and George B. Arhonditsis<sup>1\*</sup>**

<sup>1</sup> Ecological Modelling Laboratory,  
Department of Physical & Environmental Sciences, University of Toronto,  
Toronto, Ontario, Canada, M1C 1A4

\* Corresponding author

e-mail: [georgea@utsc.utoronto.ca](mailto:georgea@utsc.utoronto.ca), Tel.: +1 416 208 4858; Fax: +1 416 287 7279.

**Table 1:** Mathematical description of the plankton model: equations and their constituent processes.

$\frac{dPO_{4i}}{dt} = -a \frac{PO_{4i}}{Kp + PO_{4i}} fT_i fL_i PHYT_i P/C_{phyto}$	Phytoplankton growth
$+ \frac{\beta \lambda \left( (PHYT_i P/C_{phyto})^2 + \omega DET_i^2 \right)}{\mu^2 + (PHYT_i P/C_{phyto})^2 + \omega DET_i^2} fTz_i ZOO P_i P/C_{zoop}$	Zooplankton excretion
$+ \gamma d fTz_i ZOO P_i P/C_{zoop} \text{ or } + \gamma d fTz_i \frac{ZOO P_i^2}{pred + ZOO P_i} P/C_{zoop}$	Zooplankton mortality/ higher predation
$\text{or } + \gamma d fTz_i \frac{ZOO P_i^3}{pred^2 + ZOO P_i^2} P/C_{zoop}$	
$+ \phi fT_i DET_i$	Mineralization
$- kd (1 - \sigma_t)(PO_{4i} - PO_{4(i+1)}) \frac{A_{i+1}}{V_i (H_{i+1} - H_i)}$	Diffusion
$+ (PO_{4sed})_i \sigma_t \frac{A_i - A_{i+1}}{V_i}$	Sediment forcing
$+ PO_{4exog} *$	Exogenous loading
$+ Inflows PO_{4i} *$	Inflows
$- Outflows PO_{4i} *$	Outflows
$\frac{dPHYT_i}{dt} = a \frac{PO_{4i}}{Kp + PO_{4i}} fT_i fL_i PHYT_i$	Phytoplankton growth
$- r fT_i PHYT_i$	Phytoplankton respiration
$- \frac{\lambda (PHYT_i P/C_{phyto})^2}{\mu^2 + (PHYT_i P/C_{phyto})^2 + \omega DET_i^2} fTz_i ZOO P_i$	Herbivory
$- s (PHYT_i - PHYT_{i-1}) \frac{A_i}{V_i}$	Settling
$- Outflows PHYT_i *$	Outflows
$\frac{dZOO P_i}{dt} = \frac{\alpha \lambda \left( (PHYT_i P/C_{phyto})^2 + \omega DET_i^2 \right)}{\mu^2 + (PHYT_i P/C_{phyto})^2 + \omega DET_i^2} fTz_i ZOO P_i$	Zooplankton grazing
$- d fTz_i ZOO P_i \text{ or } - d fTz_i \frac{ZOO P_i^2}{pred + ZOO P_i} \text{ or } - d fTz_i \frac{ZOO P_i^3}{pred^2 + ZOO P_i^2}$	Zooplankton mortality/ higher predation
$- Outflows ZOO P_i *$	Outflows



$\frac{dDET_i}{dt} =$	$r fT_i PHYT_i P / C_{phyto}$	Phytoplankton respiration
	$+ \frac{\lambda[(1 - \alpha - \beta)(PHYT_i P / C_{phyto})^2 - (\alpha + \beta)\omega DET_i^2]}{\mu^2 + (PHYT_i P / C_{phyto})^2 + \omega DET_i^2} fTz_i ZOO P / C_{zoo}$	Detritivory minus zooplankton excretion
	$- \phi fT_i DET_i$	Mineralization
	$- \psi (DET_i - DET_{i-1}) \frac{A_i}{V_i}$	Settling
	$+ DET_{exog} *$	Exogenous loading
	$+ Inflows DET_i *$	Inflows
	$- Outflows DET_i *$	Outflows

---

$$fT_i = e^{-kt(Temp_i - Temp_{ref})^2}$$

$$fTz_i = e^{-ktz(Temp_i - Temp_{ref})^2}$$

$$fL_i = 2.718 FD \frac{e^{-\alpha_i} - e^{-\alpha_0}}{ke_i H_i}$$

$$FD = 0.128 \sin \left[ 2\pi \left( \frac{t}{365} + 0.8 \right) \right] + 0.5$$

$$\alpha_{0i} = \frac{I_a}{I_s} e^{-ke_i H_{1i}}$$

$$I_a = 220 \sin \left[ 2\pi \left( \frac{t}{365} - 0.32 \right) \right] + 290$$

$$\alpha_{1i} = \frac{I_a}{I_s} e^{-ke_i H_{2i}}$$

$$\sigma_t = \frac{1 - \varepsilon \cos \left[ 2\pi \left( \frac{t}{365} \right) \right]}{1 + \varepsilon}$$

$$ke_i = kb + kc PHYT_i / 50$$

$$H_i = H_{2i} - H_{1i}$$

$$i = \begin{cases} 1 & (\text{epilimnion}) \\ 2 & (\text{thermocline}) \\ 3 & (\text{hypolimnion}) \end{cases}$$

\* Exogenous loading, inflows and outflows are only considered in the epilimnion (i=1)

---

**Table 2:** Description of the parameters that were not considered during the Bayesian calibration of the eutrophication model.

<b>Parameter</b>	<b>Description</b>	<b>Unit</b>
$kt$	Effect of temperature on phytoplankton processes	$0.005\text{ }^{\circ}\text{C}^{-2}$
$ktz$	Effect of temperature on zooplankton processes	$0.006\text{ }^{\circ}\text{C}^{-2}$
$P/C_{phyto}$	Phosphorus to carbon ratio for phytoplankton	$0.015\text{ mg P}\cdot(\text{mg C})^{-1}$
$P/C_{zoop}$	Phosphorus to carbon ratio for zooplankton	$0.029\text{ mg P}\cdot(\text{mg C})^{-1}$
$Temp_{ref}$	Reference temperature	$20\text{ }^{\circ}\text{C}$
$\varepsilon$	Shape parameter for the trigonometric function $\sigma_t$	$0.85$

**Table 3:** Parameter definitions and Markov Chain Monte Carlo posterior estimates of the mean values and standard deviations of the stochastic nodes of statistical configurations that assign to the process (structural) error terms flat gamma priors and prespecified correlation (*Model 1*), and informative gamma priors coupled with fixed correlation (*Model 3*) and two models that consider linear (*Model 4*), and hyperbolic (*Model 6*) zooplankton mortality (closure) terms.

Parameter	Description	Units	Priors		Posteriors							
			Min	Max	Model1		Model3		Model4		Model6	
					Mean	SD	Mean	SD	Mean	SD	Mean	SD
<i>a</i>	Maximum phytoplankton growth rate	day <sup>-1</sup>	1.000	3.000	1.915	0.222	2.209	0.042	1.972	0.346	2.109	0.357
<i>d</i>	Zooplankton mortality rate	day <sup>-1</sup>	0.080	0.160	0.139	0.017	0.089	0.061	0.165	0.015	0.126	0.014
<i>K<sub>p</sub></i>	Half-saturation constant for <i>PO<sub>4</sub></i> uptake	mg P m <sup>-3</sup>	5.000	30.00	8.357	1.530	10.98	2.180	10.71	3.609	9.369	2.665
<i>r</i>	Phytoplankton respiration rate	day <sup>-1</sup>	0.010	0.100	0.069	0.014	0.070	0.013	0.074	0.033	0.107	0.023
<i>s</i>	Phytoplankton sinking loss rate	m day <sup>-1</sup>	0.010	0.300	0.065	0.035	0.084	0.059	0.110	0.093	0.073	0.053
<i>μ</i>	Zooplankton grazing half-saturation coefficient	mg P m <sup>-3</sup>	1.000	20.00	8.206	1.164	9.682	2.205	6.806	1.315	15.17	2.739
<i>φ</i>	Detritus mineralization rate	day <sup>-1</sup>	0.010	0.150	0.020	0.010	0.013	0.007	0.011	0.006	0.012	0.006
<i>ψ</i>	Detritus sinking rate	m day <sup>-1</sup>	0.050	1.500	0.860	0.430	0.441	0.048	0.452	0.046	0.437	0.056
<i>λ</i>	Maximum zooplankton grazing rate	day <sup>-1</sup>	0.400	0.800	0.560	0.019	0.597	0.037	0.519	0.030	0.510	0.066
<i>kb</i>	Background light extinction coefficient	m <sup>-1</sup>	0.150	0.300	0.232	0.027	0.209	0.024	0.218	0.036	0.201	0.026
<i>kc</i>	Light extinction coefficient due to chlorophyll a	m <sup>2</sup> mg <sup>-1</sup>	0.010	0.080	0.017	0.006	0.017	0.005	0.021	0.006	0.021	0.008
<i>a</i>	Zooplankton assimilation efficiency	-	0.400	0.600	0.438	0.024	0.448	0.042	0.471	0.014	0.458	0.033
<i>β</i>	Zooplankton excretion fraction to phosphate	-	0.200	0.700	0.233	0.031	0.235	0.041	0.276	0.061	0.345	0.088
<i>γ</i>	Zooplankton predation excretion fraction to phosphate	-	0.200	0.700	0.306	0.045	0.298	0.061	0.305	0.067	0.432	0.118
<i>ω</i>	Relative zooplankton preference for detritus compared to phytoplankton	-	0.200	1.000	0.224	0.047	0.159	0.044	0.219	0.106	0.222	0.072
<i>I<sub>s</sub></i>	Half saturation light intensity	MJ m <sup>-2</sup> day <sup>-1</sup>	100.0	250.0	142.3	16.93	140.4	23.09	138.0	19.93	143.7	21.81
<i>pred</i>	Half saturation constant for predation	mg C m <sup>-3</sup>	35	80	-	-	-	-	-	-	43.45	9.013











## FIGURES LEGENDS

**Figure 1:** Comparison between the prior and posterior parameter distributions of the calibration vector of the statistical configuration that assigns to the process (structural) error terms flat gamma priors and prespecified correlation (*Model 1*).

**Figure 2:** Comparison between the prior and posterior parameter distributions of the calibration vector of the statistical configuration that assigns to the process (structural) error terms informative gamma priors coupled with fixed correlation (*Model 3*).

**Figure 3:** Comparison between the prior and posterior parameter distributions of the calibration vector of the model with a linear closure term (*Model 4*).

**Figure 4:** Comparison between the prior and posterior parameter distributions of the calibration vector of the plankton model with a hyperbolic closure term (*Model 6*).

**Figure 5:** Comparison between the observed data (black dots) and the median predictions for phosphate, total phosphorus, chlorophyll *a*, and total zooplankton biomass of the statistical configurations that assign to the process (structural) error terms flat gamma priors and prespecified correlation (*Model 1*), a Wishart prior (*Model 2*), and informative gamma priors coupled with fixed correlation (*Model 3*). The 95% credible intervals represent the uncertainty pertaining to the model parameters.



**Figure 6:** Comparison between the observed data (black dots) and the median predictions for phosphate, total phosphorus, chlorophyll *a*, and total zooplankton biomass of the models with linear (*Model 4*), sigmoid (*Model 5*) and hyperbolic (*Model 6*) closure term. The 95% credible intervals represent the uncertainty pertaining to the model parameters.

**Figure 7:** Comparison between the observed phytoplankton (chlorophyll *a*) biomass (black dots) and (A) the median values of the stochastic variable “true observations”, and (B) the median predicted values of the total phytoplankton biomass, based on the models with linear (*Model 4*), sigmoid (*Model 5*) and hyperbolic (*Model 6*) closure term.

**Figure 8:** Predictions of the summer epilimnetic total phosphorus and chlorophyll *a* concentrations along with their joint posterior distribution, based on the model that assigns a Wishart prior to the process error terms (*Model 2*) and the one that considers informative gamma priors coupled with fixed correlation (*Model 3*).

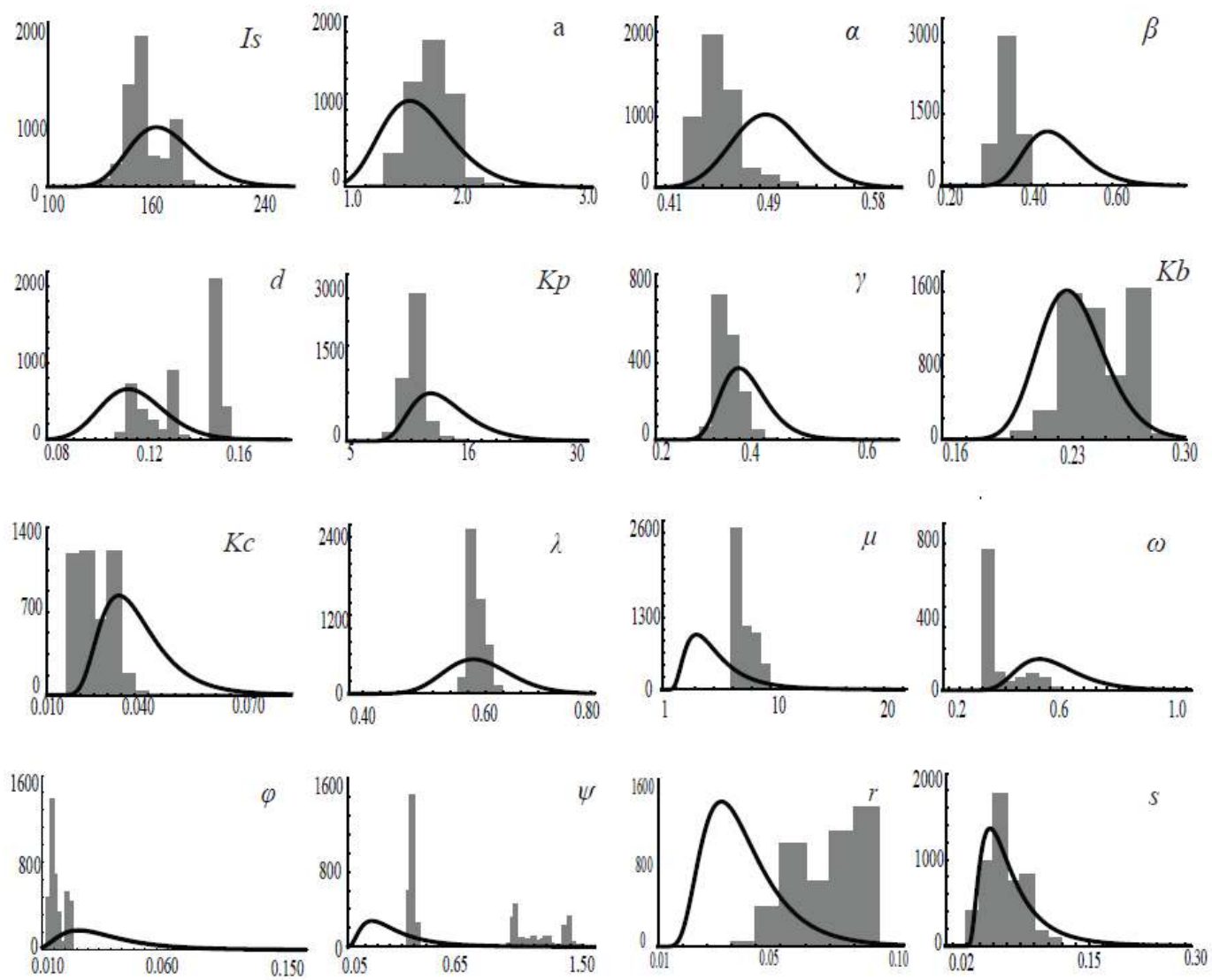


Figure 1

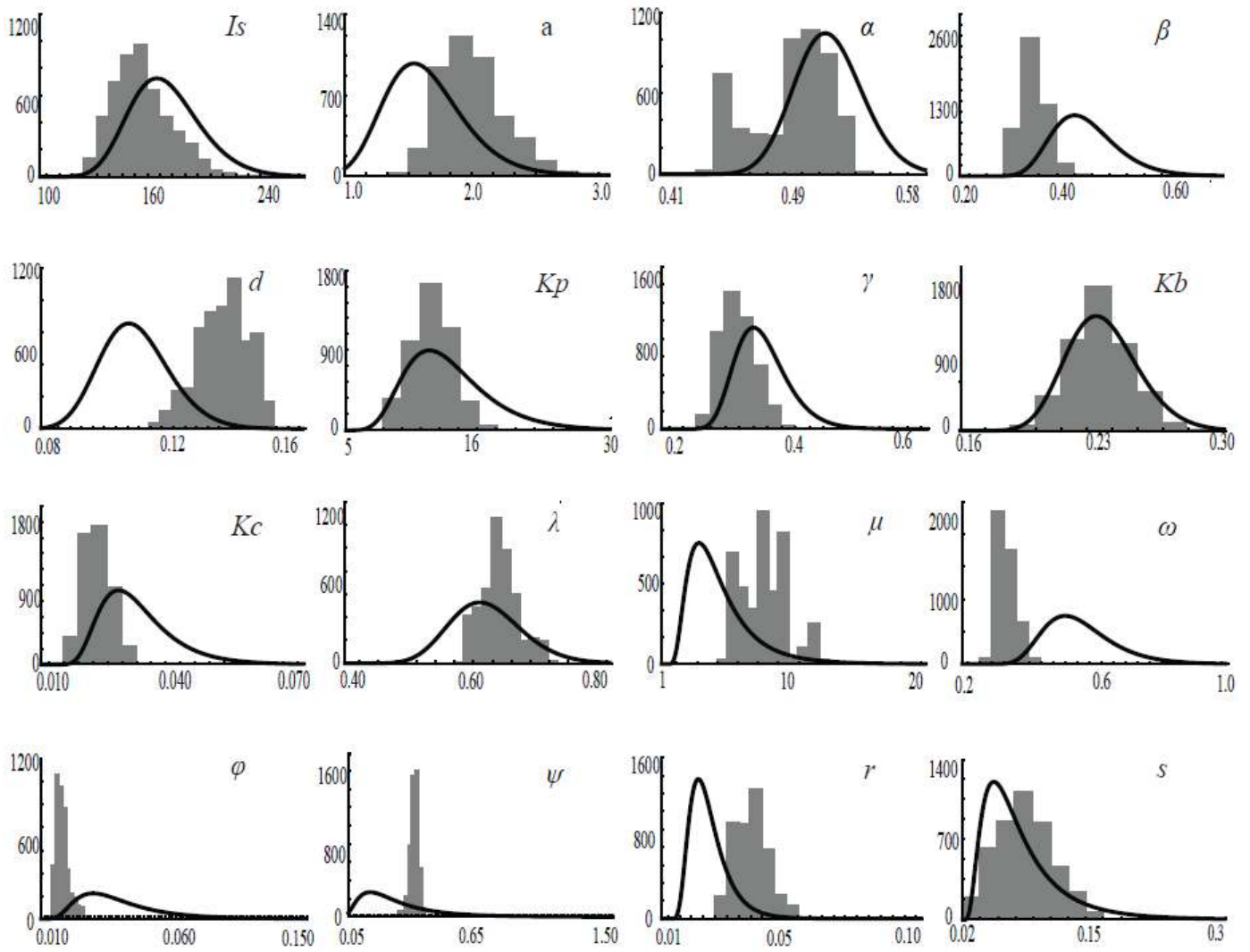


Figure 2

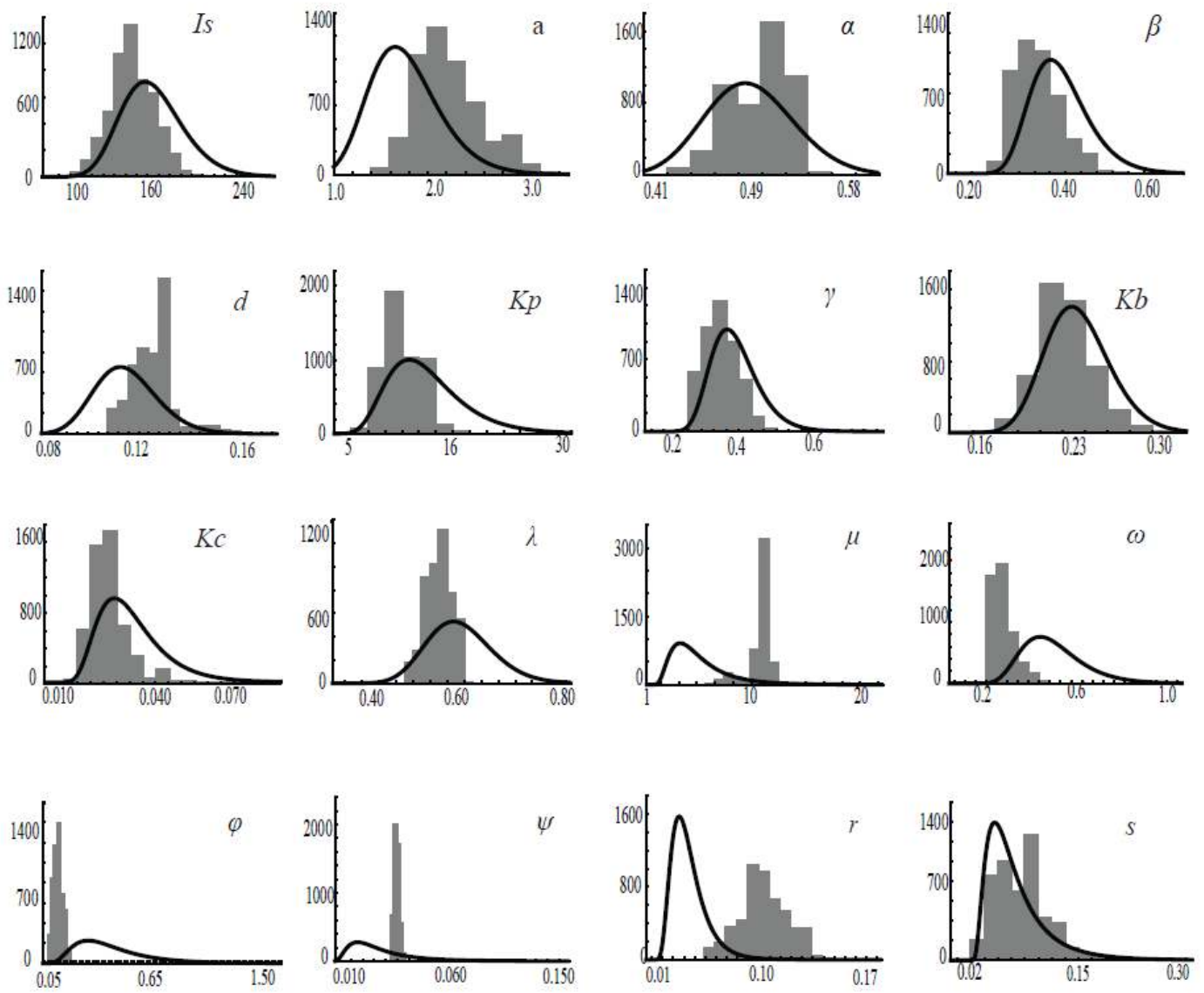


Figure 3



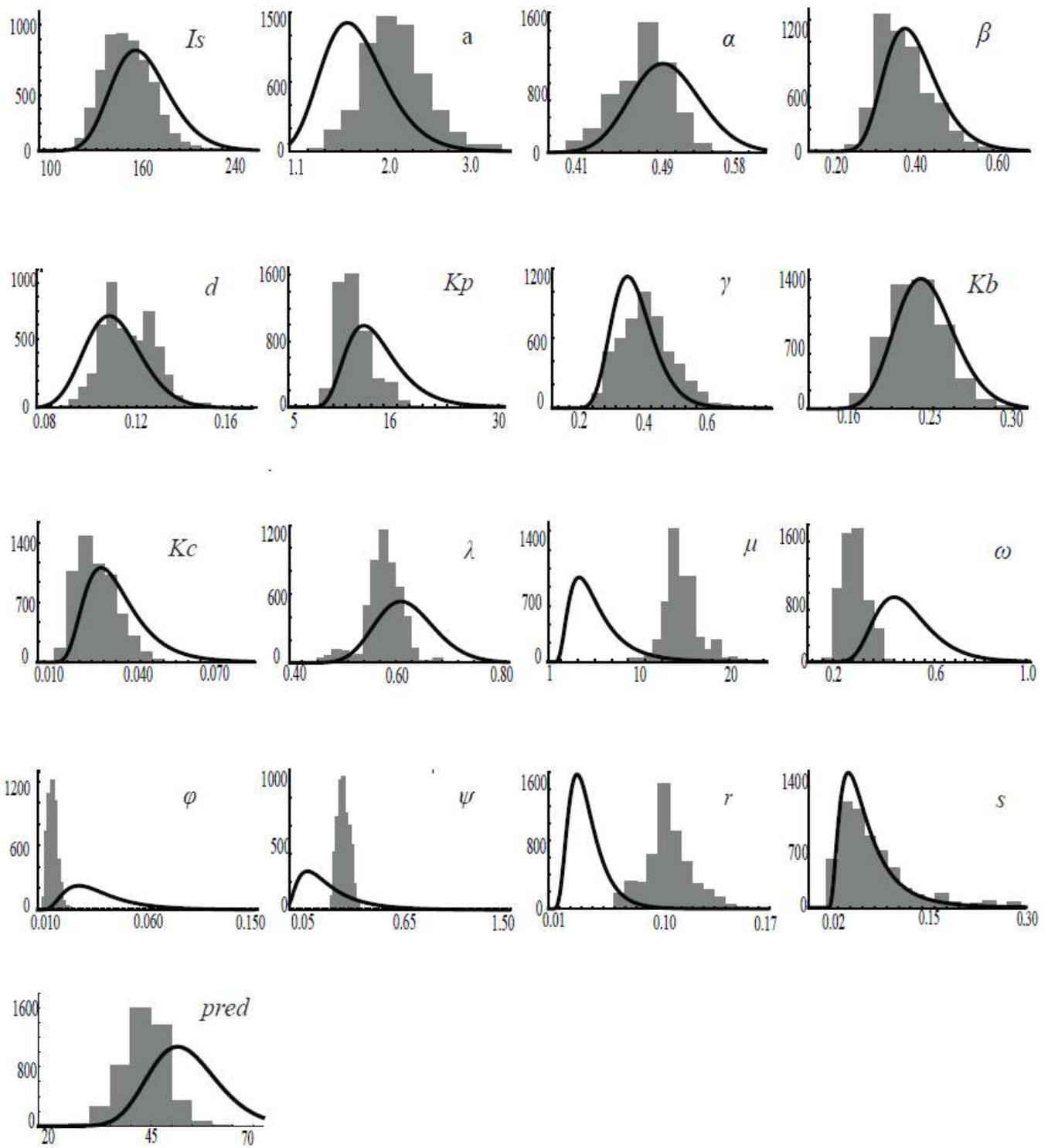
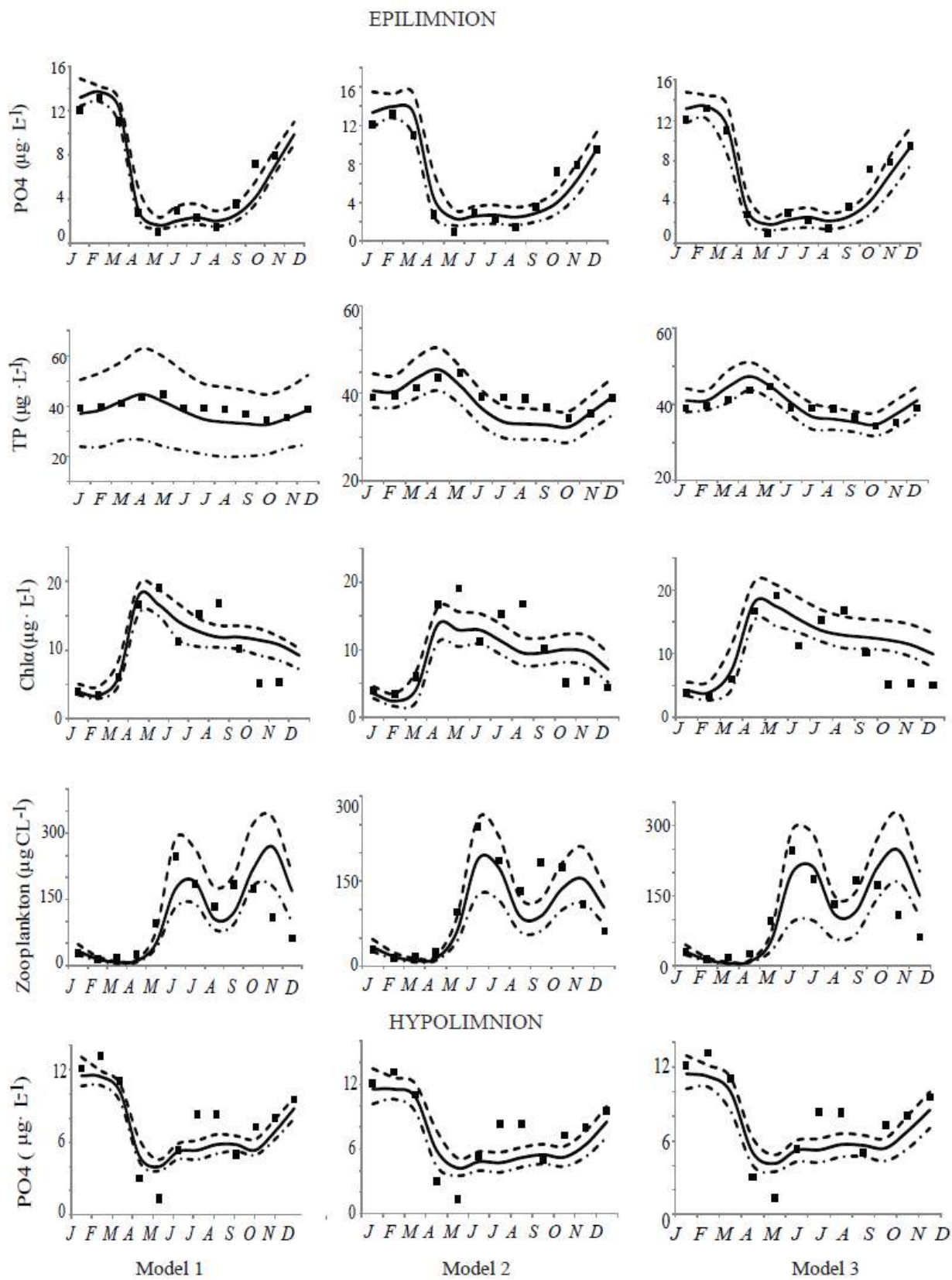
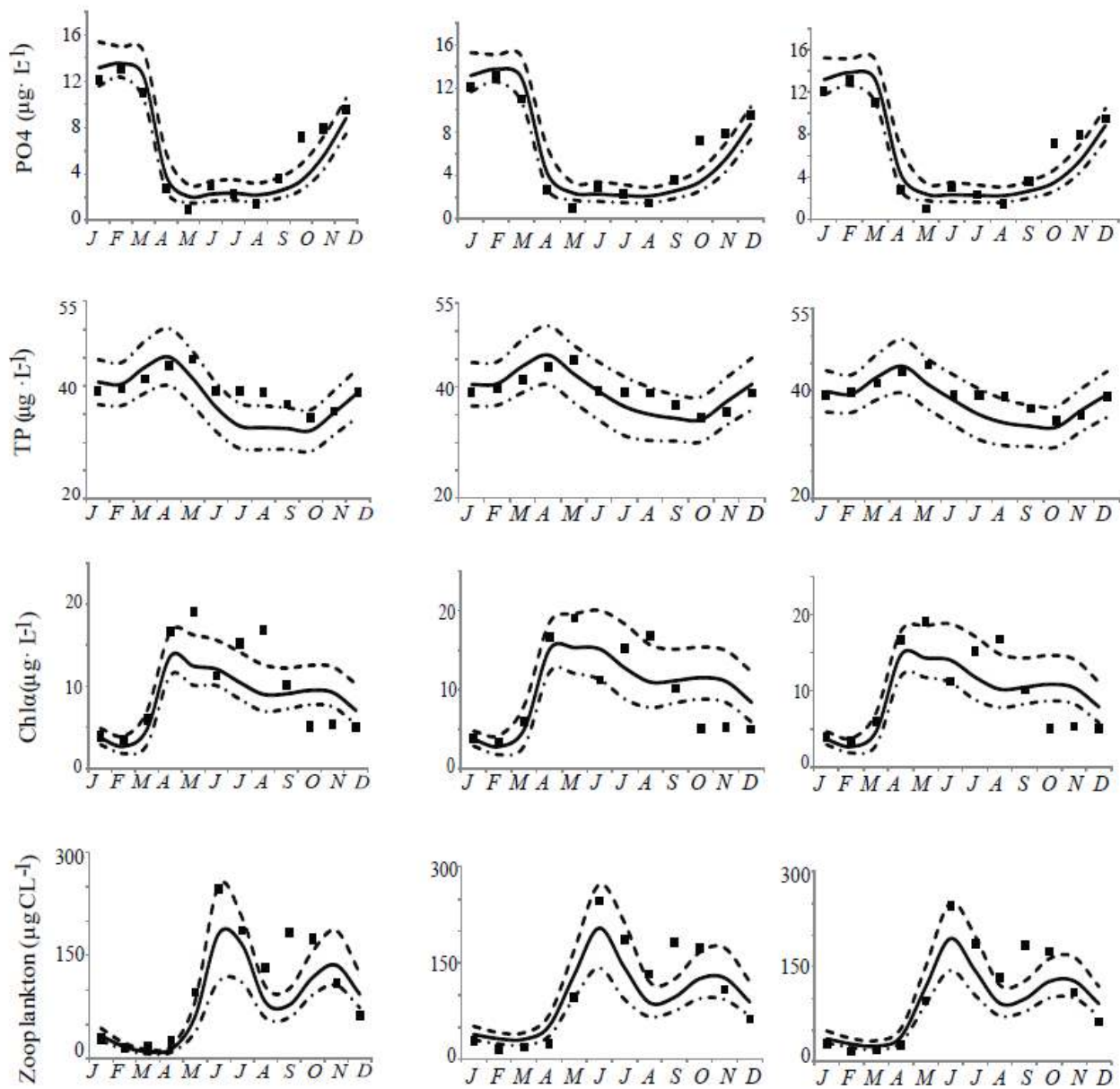


Figure 4

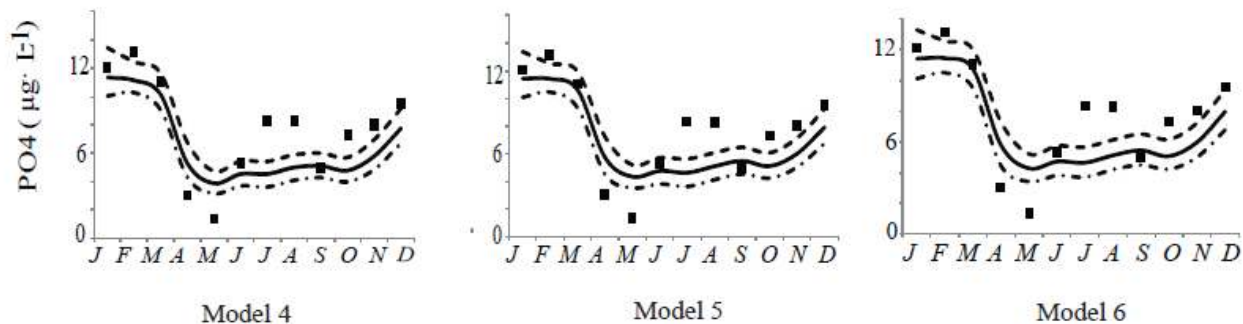


**Figure 5**

EPILIMNION



HYPOLIMNION



Model 4

Model 5

Model 6

Figure 6

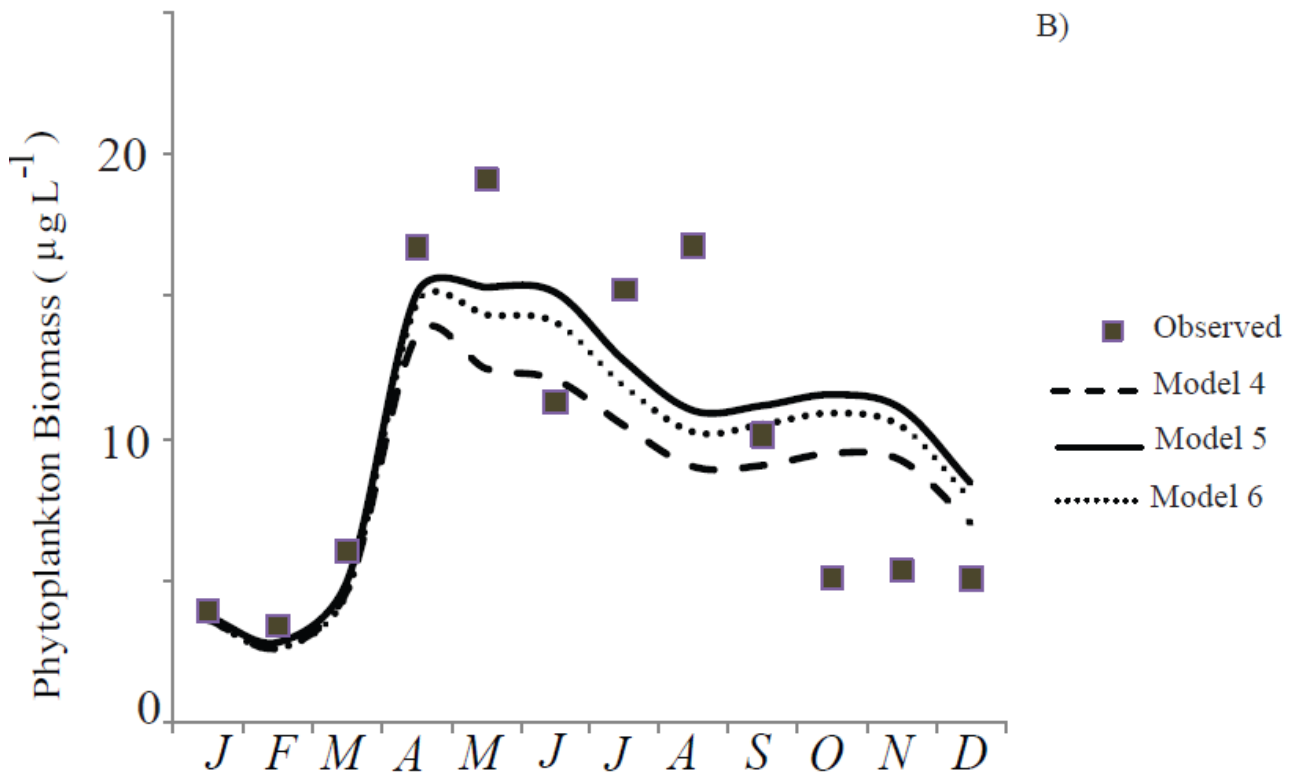
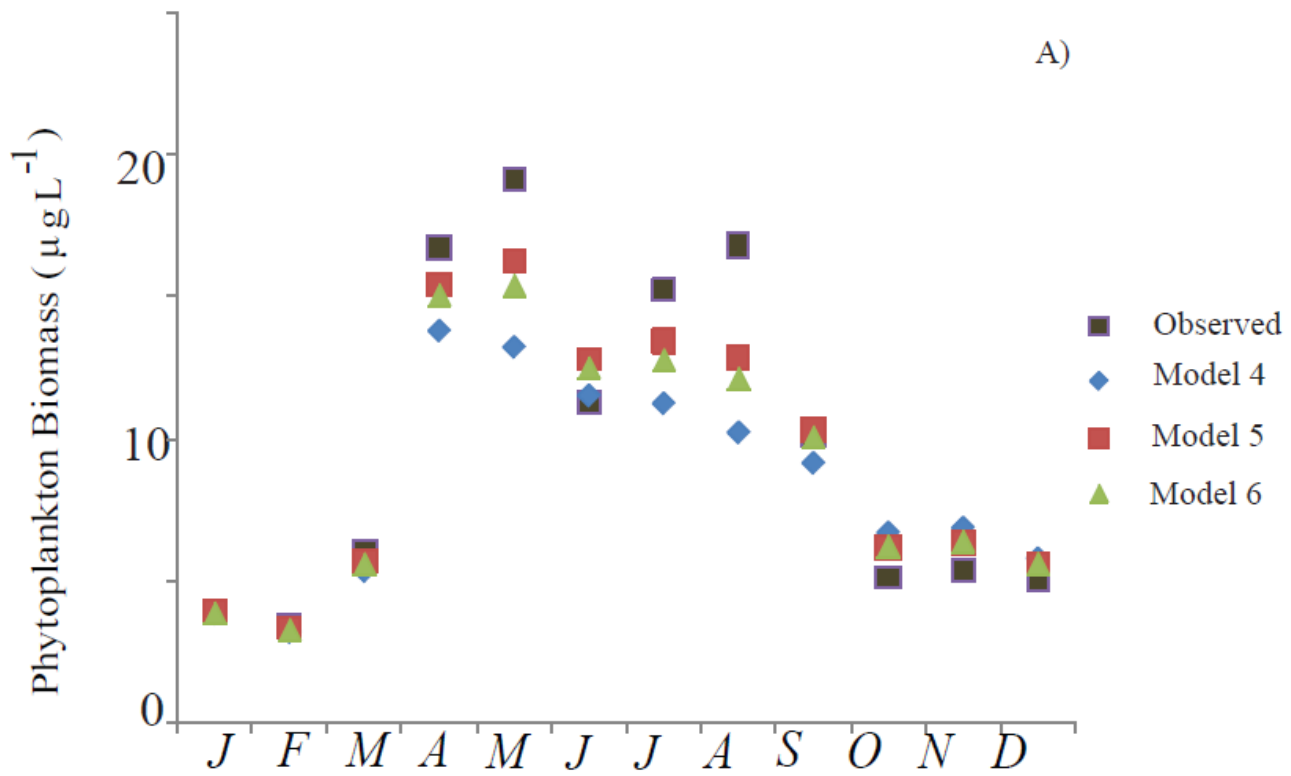
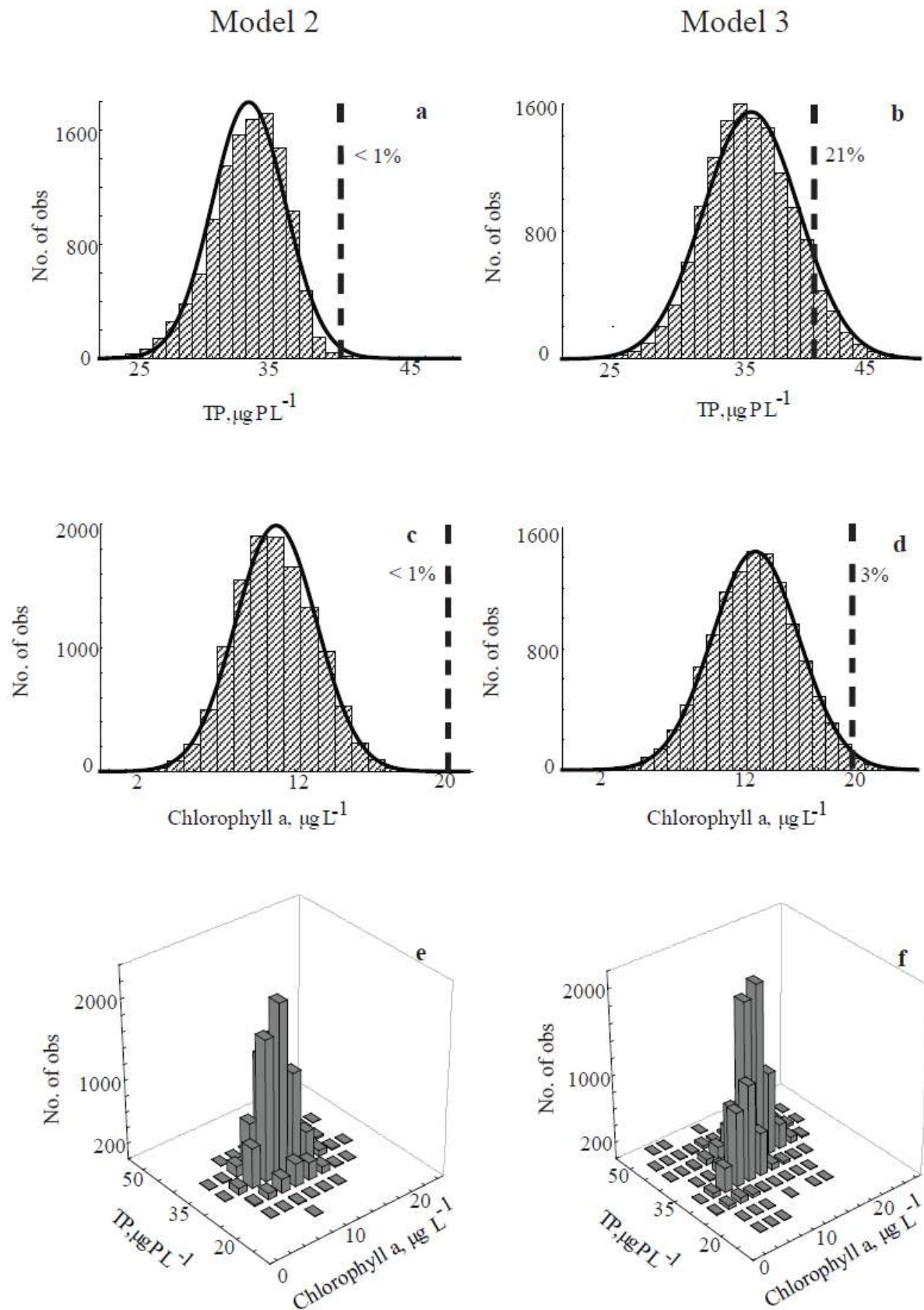


Figure 7



**Figure 8**