Ecol Res (2010) 25: 59–76
DOI 10.1007/s11284-009-0630-5

ORIGINAL ARTICLE

Vincent Cheng · George B. Arhonditsis
Michael T. Brett

A revaluation of lake-phosphorus loading models using a Bayesian hierarchical framework

Received: 25 February 2009 / Accepted: 9 June 2009 / Published online: 18 July 2009
© The Ecological Society of Japan 2009

Abstract We revisit the phosphorus-retention and nutrient-loading models in limnology using a Bayesian hierarchical framework. This methodological tool relaxes a basic assumption of regression models fitted to data sets consisting of observations from multiple systems, i.e., the systems are assumed to be identical in behavior, and therefore the models have a single common set of parameters for all systems. Under the hierarchical structure, the models are dissected into levels (hierarchies) that explicitly account for the role of significant sources of variability (e.g., morphometry, mixing regime, geographical location, land-use patterns, trophic status), thereby allowing for intersystem parameter differences. Thus, the proposed approach is a compromise between site-specific (where limited local data is a problem) and globally common (where heterogeneous systems in wide geographical areas are assumed to be identical) parameter estimates. In this study, we used critical values of the mean lake depth ($\bar{z} = 10.3$ m) and the hydraulic residence time ($\tau_w = 2.6$ years) to specify the hierarchical levels of the models. Our analysis demonstrates that the hierarchical configuration led to an improvement of the performance of six out of the seven hypothesized relationships used to predict lake-phosphorus concentrations. We also highlight the differences in the posterior moments of the group-specific parameter distributions, although the inference regarding the importance of different predictors (e.g., inflow-weighted total phosphorus input concentration, and hydraulic retention time) of lake phosphorus or the relative predictability of the models examined are not markedly different from an earlier study by Brett and Benjamin. The best fit to the observed data was obtained by the model that considers the first-order rate coefficient for total phosphorus loss from the lake as an inverse function of the lake hydraulic retention time. Finally, our analysis also demonstrates how the Bayesian hierarchical framework can be used for assessing the exceedance frequency and confidence of compliance of water-quality standards. We conclude that the proposed methodological framework will be very useful in the policy-making process and can optimize environmental management actions in space and time.

Keywords Total phosphorus · Bayesian hierarchical modeling · Hydraulic retention time · Sedimentation · Water-quality standards · Confidence of compliance

> …the efficacy of predictive limnology is not a matter of opinion. It is a matter of record…For applied limnologists, predictive limnology…has shown what sort of ecology is effective, what sort of information will sway politicians and governments to action, and how scientists can help to improve our world…
>
> R. H. Peters (1986)

Introduction

Vollenweider's (1968, 1975, 1976) research on the regulation of lake productivity by phosphorus inputs has been one of the most influential contributions in limnology and has guided much of the eutrophication research and current lake-management practices. Founded upon the "continuously stirred tank reactor principle", his steady-state, mass-balance model has provided the basis for a family of models that predict lake total phosphorus concentrations (TP_{lake}) as a function of lake morphometric/hydraulic characteristics, such as the areal phosphorus loading rate, mean lake

Electronic supplementary material The online version of this article (doi:10.1007/s11284-009-0630-5) contains supplementary material, which is available to authorized users.

V. Cheng · G. B. Arhonditsis (✉)
Department of Physical and Environmental Sciences,
University of Toronto, Toronto, ON M1C 1A4, Canada
E-mail: georgea@utsc.utoronto.ca
Tel.: +1-416-2084858
Fax: +1-416-2877279

M. T. Brett
Department of Civil and Environmental Engineering,
University of Washington, Seattle, WA 98195, USA

depth, fractional phosphorus retention, and areal hydraulic loading (Ahlgren et al. 1988; Brett and Benjamin 2008). Most of these empirical eutrophication models have been derived from cross-sectional datasets, which consist of multiple point measurements or single averages from a number of lakes. Because of the significant intersystem variability, such datasets are usually characterized by wider ranges for the variables considered in the empirical relationships and the subsequent model fit (typically using ordinary least squares) generally provides well-determined parameters (Reckhow and Chapra 1983). The resulting models are then used to predict changes within a single system at different points in time under the assumption that the large-scale (cross-sectional) patterns described in the model are also representative of the dynamics of individual systems (Reckhow 1993). By doing so, we essentially assume that all the systems in the dataset have identical behavior and therefore the empirical relationships are the same among and within lakes (Prairie and Marshall 1995). Despite their conceptual and structural simplicity, these models are viewed as important management tools for the restoration of impaired lakes and the protection of unaffected waterbodies (Dillon and Molot 1996).

While the cross-system method is a pragmatic solution to overcome the problems of overfitting when system-specific data are limited, the validity of the practice of using global-scale models to support predictions within individual lakes has been the subject of debate in the water-quality modeling literature. Earlier attempts to address this issue focused on the development of statistical techniques to detect the presence of significant effects of error variability and to unravel the "true" structural relationships (Prairie et al. 1995; Prairie and Marshall 1995). The impetus of this approach was articulated by Prairie et al. (1995) who contended that the natural error variability in many variables tends to obfuscate the delineation of existing ecological trends or to misleadingly create significant relationships where they do not actually exist. The latter problem is further accentuated by the fact that the variability within a single system is generally much smaller (and the relationships accordingly less defined) than across multiple systems. Prairie and Marshall (1995) introduced a method that addresses this problem and makes it possible to extract empirical relationships from the internal structure of a time-series within a single lake which can then be objectively compared to those obtained from cross-sectional data. Other modeling efforts have attempted to relax the assumption of globally common parameter estimates using statistical methods that allow parameter values to vary with location. A characteristic example from this modeling strategy is the development of random coefficient linear regression models to describe the nutrient-chlorophyll relationships in lakes (Reckhow 1993). In the random coefficient model, the parameters for each system are viewed as random draws from a common probability distribution (Swamy 1971; Judge et al. 1985); an assumption that can effectively accommodate the behavior of individual lakes and may result in a substantial reduction of the prediction error compared to classical (global) models (Reckhow 1993).

The random coefficient regression method is certainly an improvement over the classical approach, but it still results in point (single-valued) estimates for model parameters and therefore does not effectively depict the uncertainty associated with parameter estimates and model predictions. To overcome this problem, Borsuk et al. (2001) introduced a Bayesian hierarchical framework to model the relationship between organic matter loading and benthic oxygen demand using data from 34 estuarine and coastal systems. The basic premise of their hierarchical structure was to balance site-specific and globally common parameter estimates, i.e., each system had its own parameter set but some commonality in parameter values was assumed across systems on the basis of an underlying population distribution. With the hierarchical model configuration, the fit of the model to the observed data was improved while the adoption of the Bayesian approach allowed for a more realistic assessment of the prediction uncertainty (Borsuk et al. 2001). In particular, the Borsuk et al. (2001) study showed that the hierarchical approach usually results in lower prediction precision compared to the global model, which stems from the reduced amount of information used to estimate the site-specific parameters. The latter feature may have important implications for environmental management, although the discrepancy between the two models with regards to their predictive uncertainty can be minimized as additional data are collected for a particular system. More recently, Malve and Qian (2006) developed a Bayesian hierarchical linear model to assess compliance with chlorophyll *a* concentration standards under different nitrogen and phosphorus loads using data from the national water-quality monitoring program of Finnish lakes. The structure of their hierarchical model (i.e., number of hierarchical levels and type of groups) was based on a geomorphological classification scheme that was shown to closely represent the phytoplankton response to various nutrient levels within each lake. Aside from the higher model performance, the Malve and Qian (2006) study also underscored the ability of hierarchical modeling to transfer information across systems and support predictions in lakes with few observations and limited observational range.

In this study, we revisit the phosphorus-retention and nutrient-loading models in limnology using a Bayesian hierarchical framework. As previously discussed, this methodological approach is a compromise between site-specific (where limited local data is a problem) and globally common parameter estimates (where heterogeneous systems in wide geographical areas are assumed to be identical). Under the hierarchical structure, the models are dissected into levels (hierarchies) that explicitly account for the role of significant sources of variability, e.g., morphometry, mixing regime, geographical location, land-use patterns, and trophic status.

Our study builds upon the results of an earlier study by Brett and Benjamin (2008) and examines seven general relationships among key morphological and hydraulic lake characteristics to predict TP_{lake}. The relative performance of the seven empirical models is assessed under hierarchical and non-hierarchical formulations. Our analysis also demonstrates how the Bayesian hierarchical framework can be used to assess the exceedance frequency and confidence of compliance for different water-quality standards. Finally, we discuss ways that the proposed methodological framework can assist in the decision-making process and facilitate environmental management actions in space and time.

Methods

Dataset description

The dataset consists of 305 North American and European lakes, compiled from eight previously published, large-scale phosphorus mass-balance budgets. For each lake included in the dataset, we have the basic morphometric and hydrologic characteristics (i.e., mean lake depth, lake surface area, lake volume, inflow rate, and hydraulic retention time) along with the exogenous phosphorus loading and the in-lake TP concentration. Twenty-three lakes from the originally published phosphorus budgets are excluded from our dataset because their phosphorus loadings came from indirect estimation from land-cover data, they had unrealistic input and output phosphorus concentrations, or their TP_{lake} concentrations were suspected to be far from steady state (Brett and Benjamin 2008). Among the lakes of our dataset, there were eight Swiss lakes investigated by Vollenweider (1969); the small surface area, shallow depth (median $\bar{z} = 2.0$), low flushing rate, eutrophic Iowa lakes ($n = 16$) from the Jones and Bachman (1976) study; the large surface area, fast flushing Tennessee Valley Authority reservoirs ($n = 18$) from Higgins and Kim (1981). A large part (44%) of our database consists of the north eastern and north central United States lakes ($n = 134$) from the National Eutrophication Survey (USEPA 1975). Another large part of the database comes from the Organization for Economic Co-operation and Development (OECD) Eutrophication Programme on various settings: North American lakes ($n = 30$) (Rast and Lee 1978), European Alpine lakes ($n = 20$) (Fricker 1980), Scandinavian lakes ($n = 14$) (Ryding 1980), and the (mainly) oligotrophic Canadian Shield forest lakes ($n = 65$) from Janus and Vollenweider (1981). Summary statistics of the general limnological characteristics of the lakes included in this analysis are provided in Table 1-ESM (Electronic Supplementary Material). Lake morphometry extends over several orders of magnitude, i.e., the lake volume spans a wide range from 3.6×10^4 to 1.2×10^{12} m^3, the surface area ranges from 0.0067 to 82,367 km^2, and the mean depth varies from 0.6 to 313 m. [Note that we consider several large lakes (e.g., Great Lakes, Lake Maggiore, Lake Okanagan, Lake Seneca, Lake Tahoe, and Lake Geneva) even though they are incompletely mixed, and thus do not completely conform to the assumptions of the "continuously stirred rank reactor model"]. The dataset also covers a wide variety of trophic states, ranging from ultraoligotrophic (3 μg TP L^{-1}) to hypereutrophic (1,525 μg TP L^{-1}) systems. In particular, our dataset has 74 hypereutrophic (≥100 μg TP L^{-1}), 89 eutrophic (30–100 μg TP L^{-1}), 88 mesotrophic (10–30 μg TP L^{-1}), and 54 oligotrophic (< 10 μg TP L^{-1}) lakes (Nürnberg 1996). Aside from the Janus and Vollenweider's (1981) study, the median TP_{lake} of all the sources in the database was higher than 25 μg TP L^{-1}, while four studies had median in-lake TP concentrations equal to or higher than 50 μg TP L^{-1}.

In addition to the central tendency and dispersion measures, we also assessed the skewness (g_1) and kurtosis (g_2) of all the limnological variables. As shown in Table 1-ESM, all the limnological variables were right-skewed, i.e., the mass of the distribution is concentrated on the left and the right tail is longer. The volume and hydraulic retention are heavily right-skewed with g_1 equal to 13.4 and 15.3, respectively. The areal hydraulic loading is also strongly right-skewed ($g_1 = 10.4$) followed by the surface area ($g_1 = 9.6$) and mean depth ($g_1 = 5.0$). Interestingly, the inflow-weighted TP concentration (TP_{in}) is more heavily right-skewed compared to the in-lake total phosphorus concentration (TP_{lake}), while the ratio TP_{lake}/TP_{in} is nearly symmetrical with $g_1 = 0.78$. On the other hand, kurtosis measures the extent to which a dataset is weighted in the tails versus the center of distribution. The distributions of all the limnological variables were leptokurtic (Table 1). In particular, the hydraulic retention time ($g_2 = 250$) and the lake volume ($g_2 = 197$) were the most strongly leptokurtic variables followed by the areal hydraulic loading ($g_2 = 138$), the inflow-weighted TP concentration ($g_2 = 104$), and the surface area ($g_2 = 97$).

Empirical eutrophication models

The foundation for predicting the total phosphorus concentrations in lakes was first proposed by Vollenweider (1969) with his mass-balance model. Under steady-state conditions, this model is expressed as:

$$TP_{lake} = \frac{L}{\bar{z}(\rho + \sigma)} \tag{1}$$

This relationship is mathematically equivalent to the classic model from chemical engineering relating input and output concentrations of a substance that undergoes a first-order decay reaction in a continuous flow stirred tank reactor (Higgins and Kim 1981; Welch 1992; Brett and Benjamin 2008):

Table 1 Summary of the seven hypotheses examined in this analysis

Hypothesis	Formula	Description
H1	$\frac{TP_{in}}{1+k_1\tau_w}$	This model assumes σ is the same in all lakes, i.e., $\sigma = k_1$, where k_1 is a constant
H2	$\frac{TP_{in}}{1+k_2}$	This model considers only one predictor, TP_{in}, and implicitly assumes that $\sigma = k_2/\tau_w$, where k_2 is a constant
H3	$\frac{TP_{in}}{1+v\tau_w/\bar{z}}$	This model relates σ to the ratio of TP settling velocity (v) over the mean lake depth ($\bar{z}$), i.e., $\sigma = v/\bar{z}$.
H4	$\frac{TP_{in}}{1+k_4\tau_w^{x_4}}$	This model considers σ as an inverse function of the lake's hydraulic retention time, i.e., $\sigma = k_4\ \tau_{w4}^{x-1}$, where k_4 and x_4 are constants
H5	$\frac{a(TP_{in})}{1+b\tau_w}$	The model is given by the equation proposed by Jones and Bachman (1976) and Prairie (1988, 1989), where a and b are constants
H6	$k_6L^{x_6}$	This model assumes that TP_{lake} is proportional to L raised to some power, where k_6 and x_6 are constants
H7	$k_7\left(\frac{L}{\bar{z}}\right)^{x_7}$	This model assumes that TP_{lake} is proportional to $L/\bar{z}$ raised to some power, where k_7 and x_7 are constants

Definitions of the terminology and symbols of the seven hypotheses examined in this analysis:
TP_{lake} = TP concentration in the lake and its outflow (μg L^{-1})
TP_{in} = inflow weighted TP concentration (μg L^{-1})
L = areal TP loading rate (mg TP m^{-2} $year^{-1}$), $L = (Q \times TP_{in})/A_L$
σ = first-order rate coefficient for TP loss (or sedimentation) from the lake ($year^{-1}$)
v = settling velocity of particulate phosphorus (m $year^{-1}$)
$\bar{z}$ = mean lake depth (m), $\bar{z} = V/A_L$
A_L = lake surface area (m^2), $A_L = V/\bar{z}$
V = lake volume (m^3), $V = A_L \times \bar{z}$
ρ = flushing rate ($year^{-1}$), $\rho = 1/\tau_w = V/Q$
τ_w = mean hydraulic retention time (years), $\tau_w = 1/\rho = Q/V$
q_s = areal hydraulic loading (m $year^{-1}$), $q_s = Q/A_L = \rho \times \bar{z} = \bar{z}/\tau_w$
Q = hydraulic inflow rate (m^3 $year^{-1}$), $Q = q_s \times A_L = V/\tau_w = V \times \rho$

$$TP_{lake} = \frac{TP_{in}}{1 + \sigma\tau_w} \quad (2)$$

In this paper, we will use the latter formulation to represent the Vollenweider loading model. While many authors have used this equation, a few other formulations have been proposed in the limnological literature (Table 1). In addition, there are different interpretations of its phosphorus loss term, σ, that have generated different TP_{lake} models. Our analysis aims to examine some of these formulations and assess their relative capacity to predict TP concentrations in each lake of the database.

Hypothesis 1

$$TP_{lake} = \frac{TP_{in}}{1 + k_1\tau_w} \quad (3)$$

In this relationship, k_1 is the adjustable parameter. This model implies $k_1 = \sigma$, which means σ is the same across all lakes. Jones and Bachman (1976) found in their analysis that the best fit between predicted and observed TP_{lake} was obtained when $\sigma = 0.65$.

Hypothesis 2

$$TP_{lake} = \frac{TP_{in}}{1 + k_2} \quad (4)$$

In this model, we adjust the parameter k_2. This model is derived from the Vollenweider equation by setting $\sigma = k_2/\tau_w$ to predict TP_{lake} solely based on TP_{in}, the inflow-weighted phosphorus concentration; that is, to test how strongly inflow-weighted loading and in-lake phosphorus concentrations are related. Several authors (Schindler et al. 1978; Yeasted and Morel 1978; Reckhow 1988) have found TP_{in} to be the single best predictor of TP_{lake}. We will test to what extent this concept applies to our dataset.

Hypothesis 3

$$TP_{lake} = \frac{TP_{in}}{1 + v\tau_w/\bar{z}} \quad (5)$$

Many limnological studies (Chapra 1975; Dillon and Kirchner 1975; Kirchner and Dillon 1975; Snodgrass and O'Melia 1975; Vollenweider 1975, 1976; Larsen and Mercier 1976; Ostrofsky 1978; Higgins and Kim 1981; Nürnberg 1984, 1998; Dillon and Molot 1996) have adopted this approach, expressing σ as the ratio of the settling velocity, v, of total phosphorus to the lake mean depth, $\bar{z}$. Depending on the dataset used, the constant parameter v has been assigned a wide range of values. For example, Vollenweider (1975) found that σ approximately equals to $10/\bar{z}$, whereas Chapra (1975) fitting Kirchner and Dillon's (1975) dataset calculated $v = 16$ m $year^{-1}$. Dillon and Kirchner (1975) found that $v = 13.2$ m $year^{-1}$, which is somewhat higher than Larsen and Mercier's (1976) value of 11.73 m $year^{-1}$.

This model treats the removal of phosphorus to lake sediments using the assumption that a fraction of the TP_{lake} is in particulate form (Chapra 1975) or attached to settling particles (Brett and Benjamin 2008). However, it has an implicit (and most likely unrealistic) assumption that the phosphorus associated with particles is permanently lost once it reaches the sediment (Brett and Benjamin 2008). The latter assertion was also noted by Ahlgren et al. (1988) who concluded that v (just as σ) probably cannot be treated as a constant.

Hypothesis 4

$$TP_{lake} = \frac{TP_{in}}{1 + k_4 \tau_w^{x_4}} \tag{6}$$

This model postulates that $\sigma = k_4 \tau_{w4}^{x-1}$, where k_4 and x_4 are adjustable parameters. Many authors have found that x_4 has an approximate value of 0.5. For example, albeit the derivation was justified as being the result of "certain more or less defendable shortcuts", Vollenweider (1976) found that $\sigma = 1/\sqrt{\tau_w}$ yields the best results and adopted this approach in his model. Similarly, Larsen and Mercier (1976) derived the same expression using 20 selected lakes with $TP_{in} \leq 25\ \mu g\ TP\ L^{-1}$ to avoid internal P loading that might obscure the relationship between σ and other lake properties. Reckhow (1977), Walker (1977), Uttormark and Hutchins (1978) have also derived similar relationships which were later verified by Chapra and Reckhow (1979) with a larger database ($n = 117$).

Hypothesis 5

$$TP_{lake} = \frac{a(TP_{in})}{1 + b\tau_w} \tag{7}$$

In this model, a and b are the empirical constants. Jones and Bachman (1976) obtained the best fit to their data using an equation equivalent to the above. The appropriateness of this equation was also advocated by Prairie (1988, 1989), although in his model the coefficients a and b represent the fraction of the inflowing load that remains in the water column and the fraction of the in-lake phosphorus lost to the sediment annually, respectively.

Hypothesis 6

$$TP_{lake} = k_6 L^{x_6} \tag{8}$$

This model assumes that TP_{lake} is only related to L, and k_6, x_6 are the adjustable parameters. This hypothesis can be thought of as one form of the original Vollenweider equation in that k is equal to $1/(\bar{z}(\rho + \sigma))$. This can be viewed as how σ changes in response to $\bar{z}$ and ρ.

Hypothesis 7

$$TP_{lake} = k_7 \left(\frac{L}{\bar{z}}\right)^{x_7} \tag{9}$$

In this model, k_7 and x_7 are the adjustable parameters. This model is similar to hypothesis 6 but instead of L, TP_{lake} is predicted as a function of $L/\bar{z}$, i.e., the volumnar loading, which is identical to TP_{in}/τ_w.

Statistical analysis

Classification and regression tree analysis

We used classification and regression tree analysis (CART) to assess the role of lake morphometric and hydrologic characteristics on the TP_{lake} concentrations. Tree-based models are used in classification and regression problems when we do not want to specify a priori the form of important interactions between independent variables (Breiman et al. 1984; De'ath and Fabricius 2000). The purpose of the CART analysis is to determine a set of hierarchical decision rules (i.e., if-then split conditions) that provide optimal separation among observations (Clark and Pregibon 1992). CART models have been applied to predict abundance and composition in fish communities (Magnuson et al. 1998), to study PCB contamination in the Great Lakes (Lamon and Stow 1999; Amrhein et al. 1999), to predict dissolved oxygen levels (Nerini et al. 2000), to analyze pesticide and herbicide data (Qian and Anderson 1999), to construct regional-scale eutrophication models (Lamon and Stow 2004), and to identify the role of different functional properties and abiotic conditions on plankton community structure (Zhao et al. 2008a, 2008b). In the context of the present analysis, CART analysis was used to identify the important morphometric (volume, surface area, and mean depth) and hydrologic (inflow rates and hydraulic retention time) characteristics along with the ideal cutoff levels associated with the TP_{lake} variability. This information was then used to separate the lakes into (relatively) homogeneous subgroups, thereby dictating the optimal configuration of the Bayesian hierarchical framework.

Hierarchical Bayes

The ability of the hierarchical Bayes to decompose the environmental problems into intuitively manageable levels offers a powerful tool to disentangle complex ecological patterns, to accommodate tightly intertwined environmental processes operating at different spatiotemporal scales, to synthesize ecological information from disparate sources, and to explicitly consider the variability pertaining to latent variables or other inherently "unmeasurable" quantities (Wikle 2003a; Clark 2005). In environmental science, the hierarchical Bayes has been used to predict demographic processes and spatiotemporal population spread (Wikle 2003b; Clark 2005), to resolve the mechanisms of species coexistence and the biodiversity paradox (Clark et al. 2007), and to estimate fish population dynamics in different habitats (Wyatt 2002; Michielsens and McAllister 2004; Rivot et al. 2008).

In this study, we used a hierarchical approach to relax the assumption of globally common model parameters and therefore obtain parameter values that can (reasonably) accommodate the intersystem variability. With this approach, the problem of parameter estimation using cross-system data is viewed as a hierarchy. At the bottom of the hierarchy are the parameters θ_{ij} for individual waterbodies. At the next level, spatial heterogeneity is considered by introducing "regional" distributions; i.e., depending on the significance of various factors (morphometry, mixing regime, geographical location, land-use patterns, trophic status) the model parameters are drawn from one of these local populations θ_j. Similarly, in the upper stage, the moments of the local population parameter distributions are specified probabilistically in terms of global population parameters or hyper-parameters θ (Gelman and Hill 2007). The observed data are used to estimate the system-specific model parameters θ_{ij}, the "regional" population parameters θ_j and the hyperparameters θ. Thus, the hierarchical model dissects the problem into levels and allows intersystem parameter differences. Problems of limited local data are avoided by "borrowing strength" from other systems on the basis of the underlying population distributions. In this analysis, however, because our dataset consists of one average value for each lake, we have not considered lake-specific parameters to avoid overfitting problems. Rather, the first level of the model is based on a few (fairly) homogeneous groups as derived from the CART analysis. In this regard, the structure of our model lies between the hierarchical linear model and the non-hierarchical linear type-specific dummy variable model presented in Malve and Qian (2006).

Bayesian hierarchical models Each of the seven hypothesized relationships was evaluated using a Bayesian hierarchical approach. The hierarchical formulation is summarized as follows:

$$\log(y_{ij}) \sim N\left(f\left(\theta_j, x_{ij}\right), \tau^2\right) \tag{10}$$

$$\theta_j \sim N\left(\theta, \sigma_j^2\right) \tag{11}$$

$$\theta \sim N\left(\mu, \sigma^2\right) \tag{12}$$

$$\tau \sim U(0,100) \quad \sigma_j^2 = 10,000 \tag{13}$$

$$i = 1, \ldots, 305 \quad j = 1, 2$$

where log (y_{ij}) is the observed log TP_{lake} value from the lake *i* in the group (lake type) *j*; $f(\theta_j, x_{ij})$ is the empirical model being tested; τ^2 is the model error variance; θ_j is the group-specific parameter set; x_{ij} represents the lake-specific input variables for each TP_{lake} mathematical expression; θ corresponds to the global parameters; μ and σ^2 are the mean and variance of the global parameter distributions, respectively; σ_j^2 is the group-specific variance. A non-informative uniform prior (0, 100) was used for τ and the group-specific variance, σ_j^2, was set to an exceedingly large value, as shown in Eq. (13). [It should also be noted that the posterior results remained unaltered with the use of a diffuse gamma (conjugate) prior for the model error precision ($=1/\tau^2$).] The robustness of the posterior patterns was also examined using three different global prior distributions for the parameter vector θ. Specifically, we used (1) the posteriors from the non-hierarchical models as priors for the corresponding hierarchical models (prior$_2$); (2) the prediction intervals presented in Brett and Benjamin (2008; see their Table 5) to parameterize normal (prior$_3$) and lognormal distributions (prior$_4$). Namely, we assumed that 95% of the parameter values were lying within the reported least square estimate ± standard error intervals, following normal or lognormal distributions. The effects of the prespecified 95% probability level were further assessed by examining five additional values, i.e., 80, 68.2, 50, 34.1, and 25%. Finally, a follow-up study by Kumarappah and Arhonditsis (in preparation) presents alternative hierarchical formulations with different priors for the parameters μ, σ^2, and σ_j^2.

Bayesian non-hierarchical models To examine the advantages of the hierarchical model configuration, Bayesian non-hierarchical models were also fitted for all the eutrophication models. The Bayesian non-hierarchical models have the same implicit assumption as the classical regression models (i.e., systems have identical behavior) and therefore consider a single common set of parameters. The Bayesian non-hierarchical formulation is summarized as follows:

$$\log(y_i) \sim N\left(f(\theta, x_i), \tau^2\right) \tag{14}$$

$$\theta \sim N\left(\mu, \sigma^2\right) \tag{15}$$

$$\tau \sim U(0,100) \quad \sigma^2 = 10,000 \tag{16}$$

$$i = 1, \ldots, 305$$

where log (y_i) is the observed log TP_{lake} concentration from the lake *i*; $f(\theta, x_i)$ is the hypothesized relationship being tested; τ^2 is the model error variance; θ is the parameter vector for each empirical model; μ ($=0$) and σ^2 are the mean and variance of the parameter distributions, respectively. As shown in Eq. (16), non-informative prior distributions were used for both τ and θ (prior$_1$).

Model computations

The seven empirical models were updated with the data to obtain the posterior values for the stochastic nodes of the hierarchical (θ, θ_i and τ) and the non-hierarchical (θ and τ) formulations. 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 implemented in the WinBUGS software (Lunn et al. 2000); this algorithm is based on a symmetric normal proposal distribution, whose standard deviation is adjusted over the first 4,000 iterations such that the acceptance rate ranges between 20 and 40%. We used three chain runs of 100,000 iterations and samples were taken after the MCMC simulation converged to the true posterior distribution. Convergence was assessed using the modified Gelman–Rubin convergence statistic (Brooks and Gelman 1998). Generally, we noticed that the sequences converged very rapidly (≈1,000 iterations), and the summary statistics reported in this study were based on the last 95,000 draws by keeping every tenth iteration (thin = 10) to avoid serial correlation. The accuracy of the posterior parameter values 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 Lunn et al. 2000) for all parameters was less than 5% of the sample standard deviation.

Model comparisons

Hierarchical and non-hierarchical models are compared using the deviance information criterion (DIC); a Bayesian measure of model fit and complexity (Spiegelhalter et al. 2002). DIC is given by

$$\mathrm{DIC} = \overline{D(\theta)} + p_D \tag{17}$$

where $\overline{D(\theta)}$ is the posterior mean of the deviance, a measure of residual variance in data conditional on the parameter vector θ. The deviance is defined as $-2\log$ (likelihood) or $-2\log\,[p(y|\theta)]$; p_D is a measure of the "effective number of parameters" and corresponds to the trace of the product of Fisher's information and the posterior covariance. It is specified as the posterior mean deviance of the model $\overline{D(\theta)}$ minus the point estimate of the model deviance when using the means of the posterior parameter distributions, i.e., $p_D = \overline{D(\theta)} - D(\overline{\theta})$. Thus, this Bayesian model comparison first assesses model fit or model "adequacy" (sensu Spiegelhalter et al. 2002), $\overline{D(\theta)}$, and then penalizes complexity, p_D. A smaller DIC value indicates a "better" model. We also used the coefficient of determination (r^2) to evaluate the "mean fit" of each hypothesis (Brett and Benjamin 2008). Because the model predictions in Bayesian inference are expressed in the form of distributions, the calculation of the r^2 values was based on the posterior medians of the predicted TP_{lake} distributions.

Results

CART analysis

In the CART model, we used the log-transformed values of surface area, mean depth, volume, hydraulic retention time, areal hydraulic loading, and inflow rate to evaluate their relative importance in predicting TP_{lake} variability. During the analysis, the algorithm began with the root (or parent) node, which corresponded to the original TP_{lake} data. The data were split into increasingly homogeneous subsets with binary recursive partitioning and examination of all possible splits for each predictor variable at each node, until the Gini measure of node impurity was below a pre-specified baseline (Breiman et al. 1984). The stopping rule for the analysis was that the terminal nodes (also known as leaves in the tree analogy) should not contain more cases than 30% of the size of each class.

The final CART tree represented a hierarchical structure (shown as a dendrogram) that consisted of eight terminal nodes (Fig. 1). The first split was identified at a critical mean lake depth of 10.3 m. In the left branch, the lakes (n = 199) had a mean depth less than or equal to 10.3 m, the mean lake depth is again the important predictor and a value of 1.65 m divides the lakes into two subgroups. The lakes (n = 30) with mean depth less than or equal to 1.65 m make up the first terminal node, I, characterized by the highest lake phosphorus concentration with a median of 160 μg TP L^{-1}. On the other hand, the splitting of the deeper lakes (> 10.3 m; n = 106) located in the right branch of the tree occurred at a hydraulic retention level of 17.8 years. Lakes with hydraulic retention times τ_w > 17.8 years form the last terminal node, VIII, which has the lowest TP_{lake} with a median of 8.8 μg TP L^{-1}. The group of lakes with hydraulic retention ≤ 17.8 years was further partitioned into the terminal nodes, VI and VII, based on another critical hydraulic retention time of 9.8 years. Interestingly, the surface area was an important predictor variable for lakes (n = 169) within the 1.67–10.3 m depth range. In this portion of the tree, two splits occurred at surface-area values of 1.47 km^2 and then again at 20.3 km^2. Lakes with a surface area smaller than 1.47 km^2 and larger than 20.3 km^2 formed the terminal nodes II and V, whereas lakes within this surface area range were further split into the terminal nodes III and IV based on a hydraulic retention time value of 0.17 years. Overall, the left branch of the regression tree was dominated by morphometric predictors, whereas the right branch was dominated by hydrologic predictors. While insightful as an exploratory analysis, the structure of the dendrogram in Fig. 1 was deemed quite complex for the present illustration. Therefore, for the sake of simplicity, the hierarchical model presented herein was just founded upon the first splitting condition

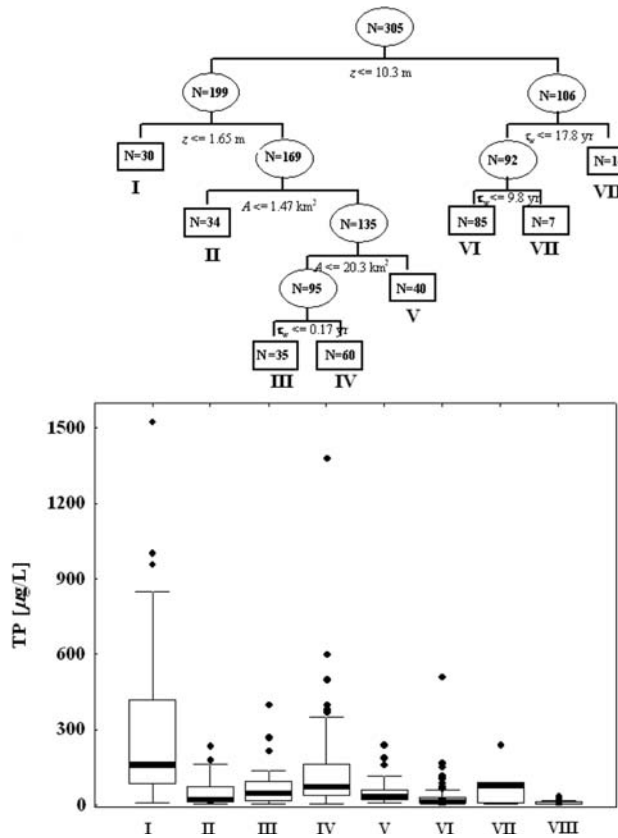


Fig. 1 Classification and regression tree diagram of observed TP_{lake} ($\mu g\ L^{-1}$) partitioned with mean depth $\bar{z}$ (m), surface area A_L (km^2), and hydraulic retention τ_w (years) along with the box-plots of TP_{lake} on each terminal node. Terminal nodes (I to VIII) are shown in *thick black rectangles*; lakes in the parent nodes are sent to the left child nodes if the corresponding values are no greater than the split conditions; otherwise they are sent to the right child nodes

($\bar{z} = 10.3$ m) that delineated two subgroups, i.e., shallow and deep lakes (“mean depth” hierarchical model). To examine the sensitivity of our results to this lake grouping, we also considered an alternative hierarchical model configuration based on a critical hydraulic retention time value (τ_w) of 2.6 years (“hydraulic retention time” hierarchical model). The latter condition was derived from a CART model that only examined the effects of the hydrologic characteristics (i.e., inflow rate, hydraulic retention, and areal hydraulic loading rate) on TP_{lake} variability.

Model fit

Comparison between hierarchical and non-hierarchical models

The model fit assessment (deviance and DIC) for the Bayesian non-hierarchical and hierarchical models is presented in Table 2. Our results show that the three priors ($prior_1$, $prior_2$, $prior_3$) resulted in nearly identical model performance, indicating that the posterior patterns were insensitive to the prior parameter distributions assigned. Generally, the hierarchical configuration improved the performance relative to the non-hierarchical approach, although the degree of improvement varied among the different hypotheses. In particular, aside from hypotheses H4 and H7, the distinction between shallow and deep lakes resulted in substantially lower deviance and DIC values. In a similar manner, the delineation of groups based on the hydraulic retention time improved the performance of four models (H1, H2, H5, and H7), whereas hypotheses H3, H4 and H6 were nearly unaltered with regards to the model fit. The comparison between the hierarchical and non-hierarchical models of the seven TP_{lake} relationships can also be illustrated by examining the posterior distributions for the model standard error terms, where the model improvement is manifested as a shift towards smaller values, i.e., hypotheses 1, 2, 3, 5, and 6 (Fig. 1-ESM). On the other hand, the error distributions associated with the hypotheses 4 and 7, which did not improve under the hierarchical model configuration, were almost identical. Similar inferences can be drawn for the hypotheses 3, 4, and 6 when using the “hydraulic retention time” model (Fig. 2-ESM).

Marginal posterior parameter distributions

The moments of the marginal posterior parameter distributions for each hypothesis under the non-hierarchical and hierarchical (mean depth and hydraulic retention) model configurations are presented in Table 3. The posterior parameter values for the Bayesian non-hierarchical models were equal to the classical least-squares estimates of the Brett and Benjamin (2008) study. This result is not surprising as the Bayesian approach combined with non-informative priors usually provides similar results to the ones obtained from the classical frequentist statistical practice (Ellison 1996). The posterior moments of the global parameter distributions of the hierarchical models were also relatively similar. The difference lies in the posterior group-specific parameters (θ_1, θ_2), which the dataset is directly fitted to and hence the differences in predictions. The comparison of the posterior group-specific parameter distributions obtained from the Bayesian hierarchical model based on the mean depth partitioning and the normally distributed global priors ($prior_3$) is shown in Fig. 2. Due to the reduced information available for estimating the group-specific parameters, the respective distributions are flatter (less precise) than the values obtained from the non-hierarchical model. Generally, the marginal group-specific parameter distributions drifted away from the global prior distribution in opposite directions, whereas notable exceptions were the parameters x_4 (H4) and x_7 (H7) with relatively unaltered locations. Importantly, the group-specific posterior patterns demonstrate

Table 2 The deviance [−2log (likelihood)] and the deviance information criterion (DIC) for the non-hierarchical and hierarchical models

	Non-hierarchical		Hierarchical											
	Prior$_1$		Mean depth $\bar{z} = 10.3$ m						Hydraulic retention $\tau_w = 2.6$ years					
			Prior$_2$		Prior$_3$		Prior$_4$		Prior$_2$		Prior$_3$		Prior$_4$	
	Deviance	DIC	Deviance	DIC	Deviance	DIC	Deviance	DIC	Deviance	DIC	Deviance	DIC	Deviance	DIC
H1	584.3	586.3	561.1	564.0	561.1	564.1	561.1	564.0	**545.7**	**548.6**	545.7	548.7	545.7	548.6
H2	654.9	656.9	616.7	619.7	616.7	619.7	616.7	619.7	544.6	547.6	**544.6**	**547.6**	544.6	547.6
H3	586.1	588.1	539.9	542.9	**539.9**	**542.9**	539.9	542.9	588.2	591.2	588.2	591.2	588.2	591.2
H4	**466.7**	**469.7**	471.2	476.1	471.2	476.2	471.2	476.2	468.8	473.7	468.7	473.7	468.7	473.7
H5	511.5	514.5	481.3	486.2	481.3	486.2	481.3	486.2	**480.7**	**485.5**	**480.7**	**485.5**	**480.7**	**485.5**
H6	867.9	870.7	828.7	832.8	**828.7**	**832.8**	828.9	832.9	870.3	873.4	870.4	873.3	870.4	873.4
H7	821.6	824.5	824.3	829.1	824.2	829.1	823.6	828.4	803.7	808.5	**803.7**	**808.5**	803.7	808.6

Prior$_1$ denotes non-informative prior parameter distributions, prior$_2$ denotes priors based on the posterior parameter distributions of the non-hierarchical models, prior$_3$ and prior$_4$ denote priors normally and log-normally distributed within the predictive intervals reported in the Brett and Benjamin (2008) study, respectively. Performance criteria values in bold font denote the best performing model for each hypothesis

Table 3 Summary of the posterior parameter distributions for each hypothesis under the non-hierarchical and hierarchical (mean depth and hydraulic retention) model configuration

Non-hierarchical models				Hierarchical models				
					Mean depth $\bar{z} = 10.3m$		Hydraulic retention $\tau_w = 2.6$ years	
Hypothesis		Median	SD		Mean	SD	Mean	SD
1	k_1 (year^{-1})	0.45	0.04	k_1	0.45	0.02	0.45	0.02
				k_{11}	0.78	0.10	1.01	0.12
				k_{12}	0.32	0.04	0.33	0.03
2	k_2	1.06	0.08	k_2	1.06	0.04	1.06	0.04
				k_{21}	0.72	0.08	0.64	0.06
				k_{22}	1.90	0.18	3.07	0.28
3	v (m year^{-1})	5.10	0.44	v	5.10	0.30	5.10	0.31
				v_1	3.20	0.38	5.32	0.74
				v_2	10.09	1.04	5.00	0.55
4	k_4 (year$^{-0.47}$)	1.12	0.08	k_4	1.12	0.04	1.12	0.04
				k_{41}	1.12	0.10	1.03	0.10
				k_{42}	1.14	0.13	1.15	0.20
	x_4	0.47	0.04	x_4	0.47	0.02	0.47	0.04
				x_{41}	0.51	0.07	0.37	0.07
				x_{42}	0.46	0.05	0.48	0.07
5	a	0.66	0.03	a	0.65	0.02	0.65	0.02
				a_1	0.77	0.04	0.75	0.04
				a_2	0.50	0.04	0.35	0.04
	b (year^{-1})	0.17	0.03	b	0.17	0.02	0.17	0.02
				b_1	0.44	0.08	0.47	0.11
				b_2	0.07	0.02	0.04	0.01
6	k_6 (mg$^{0.61}$ year$^{0.39}$ m$^{-2.22}$)	2.30	0.48	k_6	2.20	0.26	2.20	0.26
				k_{61}	4.83	1.16	2.21	0.64
				k_{62}	1.51	0.59	1.65	1.06
	x_6	0.39	0.03	x_6	0.40	0.02	0.40	0.02
				x_{61}	0.33	0.03	0.39	0.03
				x_{62}	0.39	0.06	0.48	0.10
7	k_7 (mg$^{0.66}$ year$^{0.34}$ m$^{-1.98}$)	6.61	0.76	k_7	6.50	0.36	6.50	0.35
				k_{71}	8.47	1.51	5.21	0.85
				k_{72}	5.93	1.07	4.66	0.99
	x_7	0.34	0.02	x_7	0.34	0.01	0.34	0.01
				x_{71}	0.31	0.03	0.36	0.02
				x_{72}	0.35	0.05	0.60	0.08

For the hierarchical approach, global posteriors are followed by group-specific parameter values remarkable stability relative to the precision assigned to the global priors (Fig. 3-ESM).

Hypotheses comparison

Hypothesis 1 The hypothesis that the rate coefficient for TP loss (σ) is constant led to a reasonably good model fit relative to other hypotheses with a DIC value of 586.3 and $k_1 = 0.45 \pm 0.04$ year^{-1}. Under the "mean depth" hierarchical model, the DIC was decreased to 564.1 with values of $k_{11} = 0.78 \pm 0.10$ year^{-1} and $k_{12} = 0.32 \pm 0.04$ year^{-1} for the shallow ($\bar{z} \leq 10.3$ m) and the deep lakes ($\bar{z} > 10.3$ m), respectively (Tables 2, 3; Fig. 2). The comparison between the predicted and observed TP values yielded a nonlinear r^2 of 0.857 (Fig. 4-ESM). Under the "hydraulic retention time" hierarchical model, the DIC also improved to 548.6 with $k_{11} = 1.01 \pm 0.12$ year^{-1} for the $\tau_w \leq 2.6$ years group and $k_{12} = 0.33 \pm 0.03$ year^{-1} for the $\tau_w > 2.6$ years group (Tables 2, 3; Fig. 5-ESM).

Hypothesis 2 The hypothesis that TP_{lake} is directly proportional to TP_{in} (in that the product $\sigma\tau_w$ is a constant) led to a DIC of 659.6 with $k_2 = 1.06 \pm 0.08$. The "mean depth" hierarchical model resulted in a reduced DIC value (=619.7) with $k_{21} = 0.72 \pm 0.08$ and $k_{22} = 1.90 \pm 0.18$ for the two lake groups. The comparison between the predicted and observed TP values yielded a nonlinear r^2 of 0.828. Under the "hydraulic retention time" hierarchical model, the DIC improved to 547.6 with $k_{21} = 0.64 \pm 0.06$ for the $\tau_w \leq 2.6$ years group and $k_{22} = 3.07 \pm 0.28$ for the one consisting of lakes with $\tau_w > 2.6$ years.

Hypothesis 3 The hypothesis that the rate coefficient for TP loss can be approximated as the ratio of the apparent TP settling velocity to the mean lake depth resulted in a fit slightly worse than the one obtained from the first hypothesis (H1), i.e., DIC = 588.1 with $v = 5.1 \pm 0.30$ m year^{-1}. With the "mean depth" hierarchical model, the DIC decreased to 542.9 with $v_1 = 3.20 \pm 0.44$ m year^{-1} and $v_2 = 10.09 \pm 1.04$ m year^{-1} for the shallow ($\bar{z} \leq 10.3$ m) and deep lakes ($\bar{z} > 10.3$ m), respectively. The comparison between the predicted and observed TP values gave a nonlinear r^2 of 0.866. With the "hydraulic retention time" hierarchical model, the DIC increased to 591.2 with $v_1 = 5.32 \pm 0.74$ m year^{-1} for the $\tau_w \leq 2.6$ years group and $v_2 = 5.00 \pm 0.55$ m year^{-1} for the $\tau_w > 2.6$ years group.

Hypothesis 4 The hypothesis that the rate coefficient for TP loss can be represented by an expression of the form $\sigma = k_4\tau_{w4}^{x-1}$ led to the best fit among all the hypotheses examined with the lowest DIC (=469.7) value and $k_4 = 1.12 \pm 0.08$ year$^{-0.47}$, $x_4 = 0.47 \pm 0.04$. The hierarchical configuration of the model based on the mean depth partitioning did not improve the performance, as DIC increased slightly to 476.2. The group-specific posterior parameter values were relatively similar to the corresponding global priors, i.e., $k_{41} = 1.12 \pm 0.10$ year$^{-0.51}$, $x_{41} = 0.51 \pm 0.07$ for the shallow lakes, and $k_{42} = 1.14 \pm 0.13$ year$^{-0.46}$, $x_{42} = 0.46 \pm 0.05$ for deep lakes. The comparison between the predicted and observed TP values provided a nonlinear r^2 of 0.895. With the "hydraulic retention time" hierarchical model, the DIC increased slightly to 473.7 with $k_{41} = 1.03 \pm 0.10$ year$^{-0.37}$, $x_{41} = 0.37 \pm 0.07$ for the $\tau_w \leq 2.6$ years lake group and $k_{42} = 1.15 \pm 0.20$ year$^{-0.48}$, $x_{42} = 0.48 \pm 0.07$ for the one consisting of lakes with $\tau_w > 2.6$ years.

Hypothesis 5 The posterior distribution of $a = 0.66 \pm 0.03$ and $b = 0.17 \pm 0.03$ year^{-1} was associated with a DIC equal to 514.5, which is the second best value among all the hypotheses. Under the "mean depth" hierarchical model, the DIC value decreased to 486.2 with $a = 0.77 \pm 0.04$, $b = 0.44 \pm 0.08$ year^{-1} for the $\bar{z} \leq 10.3$ m group and $a = 0.50 \pm 0.04$, $b = 0.07 \pm 0.02$ year^{-1} for the $\bar{z} > 10.3$ m group. The comparison between the predicted and observed TP values yielded a nonlinear r^2 of 0.891. Using the "hydraulic retention time" hierarchical model, the DIC decreased to 485.5 with $a_1 = 0.75 \pm 0.04$, $b_1 = 0.47 \pm 0.11$ year^{-1} for lakes with $\tau_w \leq 2.6$ years, and $a_2 = 0.35 \pm 0.04$, $b_2 = 0.04 \pm 0.01$ year^{-1} for the ones with $\tau_w > 2.6$ years.

Hypothesis 6 The hypothesis that TP_{lake} is proportional to the areal TP loading performed the poorest among all the empirical models. The posterior parameter distributions of $k_6 = 2.30 \pm 0.48$ mg$^{0.61}$ year$^{0.39}$ m$^{-2.22}$ and $x_6 = 0.39 \pm 0.03$ yielded a DIC of 870.7. With the "mean depth" hierarchical model, the DIC decreased to 832.8 and the non-linear r^2 was equal to 0.660. The posterior parameter distributions for the shallow and deep lakes were $k_{61} = 4.83 \pm 1.16$ mg$^{0.67}$ year$^{0.33}$ m$^{-2.34}$, $x_{61} = 0.33 \pm 0.03$ and $k_{62} = 1.51 \pm 0.59$ mg$^{0.61}$ year$^{0.39}$ m$^{-2.22}$, $x_{62} = 0.39 \pm 0.06$, respectively. Under the "hydraulic retention time" hierarchical model, the DIC increased slightly to 873.4 with $k_{61} = 2.21 \pm 0.64$ mg$^{0.61}$ year$^{0.39}$ m$^{-2.22}$, $x_{61} = 0.39 \pm 0.03$ for the $\tau_w \leq 2.6$ years group and $k_{62} = 1.65 \pm 1.06$ mg$^{0.52}$ year$^{0.48}$ m$^{-2.04}$, $x_{62} = 0.48 \pm 0.10$ for the $\tau_w > 2.6$ years one. [The non-intuitive units for the value of k_6 arise because, when one raises L to an empirical exponent, the units of L are raised to that exponent along with its numerical value; the units of k_6, combined with those of L^{x6}, generate the desired units for TP_{lake}.]

Hypothesis 7 The hypothesis that TP_{lake} is proportional to the ratio of areal TP loading with the mean lake depth fits the data slightly better than the sixth hypothesis with a DIC value of 824.5 for $k_7 = 6.61 \pm 0.76$ mg$^{0.66}$ year$^{0.34}$ m$^{-1.98}$ and $x_7 = 0.34 \pm 0.02$. Under the "mean depth" hierarchical model, the fit did not improve, resulting in a DIC value equal to 829.1 and a non-linear r^2 of 0.665. The posterior parameter distributions for the shallow and deep lakes were $k_{71} = 8.47 \pm 1.51$ mg$^{0.69}$ year$^{0.31}$ m$^{-2.07}$, $x_{71} = 0.31 \pm 0.03$ and $k_{72} = 5.93 \pm 1.07$ mg$^{0.65}$ year$^{0.35}$ m$^{-1.95}$, $x_{72} = 0.35 \pm 0.05$, respectively. Under the

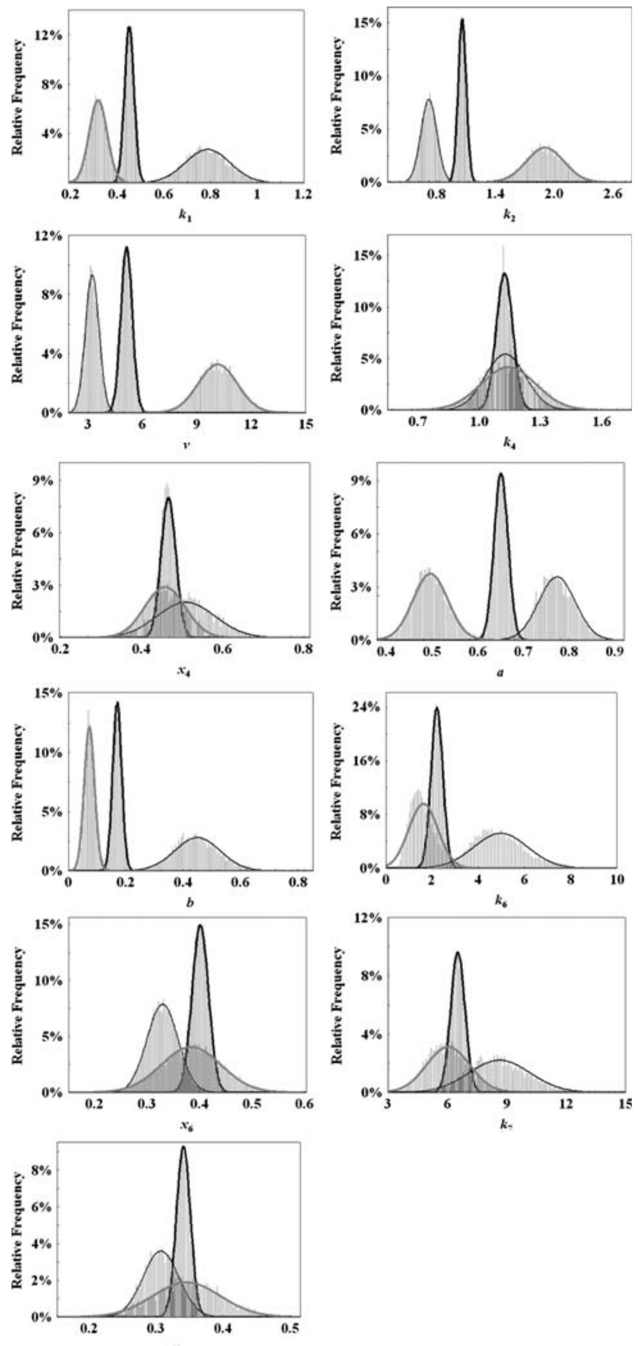


Fig. 2 Bayesian hierarchical model based on the mean depth partitioning. Prior (*thick black lines*) and posterior group-specific (*thin black* $\leq$ 10.3 m, and *thick gray* $>$ 10.3 m) parameter distributions. Global priors are normally distributed within the confidence intervals reported in the Brett and Benjamin (2008) study (prior$_3$). The sensitivity of the posterior patterns on the specification of the global prior distributions is presented in Fig. 3-ESM

"hydraulic retention time" hierarchical model, the DIC decreased to 808.5 with k_{71} = 5.21 ± 0.85 mg$^{0.64}$ year$^{0.36}$ m$^{-1.92}$, x_{71} = 0.36 ± 0.02 for the $\tau_w \leq 2.6$ years group and k_{72} = 4.66 ± 0.99 mg$^{0.40}$ year$^{0.60}$ m$^{-1.2}$, x_{72} = 0.60 ± 0.08 for lakes with $\tau_w > 2.6$ years.

Exceedance frequency and confidence of compliance with total phosphorus standards

The MCMC posterior samples were also used to examine the exceedance frequency and confidence of compliance with a total phosphorus standard in Lake Dudley (Ontario) and the Hiwassee Reservoir (Tennessee Valley). We used the highest performing hypothesis, H4, under the "mean depth" hierarchical model and specified the TP_{lake} threshold value (numerical criterion) at 25 μg TP L^{-1}. For each iteration, we calculated a predicted value and a corresponding probability of exceeding the criterion. The latter probability was calculated as follows:

$$p_{ij} = P\left(\mathrm{TP}_{ij} > \mathrm{TP}^* | \theta_j, \theta, x_{ij}, \tau\right) = 1 - F\left(\frac{\mathrm{TP}^* - f\left(\theta_j, x_{ij}\right)}{\tau}\right) \tag{18}$$

where p_{ij} is the system-specific probability of total phosphorus exceeding the numerical criterion TP* given values of θ, θ_j, and x_{ij}, τ is the model standard error, and $F(\cdot)$ is the value of the cumulative standard normal distribution. The distribution of the exceedance probability p_{ij} across the posterior space (12,500 MCMC samples) can then be used to assess the expected exceedance $\bar{p}_{ij}$ and the confidence of compliance (CC), while accounting for the uncertainty in model predictions that stems from the model parameters. Confidence of compliance (CC) is the proportion of the exceedance probability p distribution that lies below the EPA's 10% guideline (Borsuk et al. 2002). In our example, the distributions of the predicted TP_{lake} concentration in the two freshwater systems under their current, 10, 25, and 50% reduction of the TP_{in} concentrations along with the 25 μg TP L^{-1} numerical criterion are shown in Fig. 3. The corresponding expected exceedance and confidence of compliance (the proportion of the exceedance frequency distribution that lies below the EPA's 10% guideline; CC) for Lake Dudley are approximately 17.1 and 0.11%, respectively. The corresponding values in the Hiwassee Reservoir are 12.8 and 9.5% (Figs. 4, 5). This probabilistic assessment of the total phosphorus concentrations should make model results very useful to decision-makers and stakeholders, because the deterministic statements are avoided and the optimal management schemes (e.g., reduction of nutrient loading) are determined by explicitly acknowledging an inevitable risk of non-attainment. In Lake Dudley, the exceedance frequency of the 25 μg TP L^{-1} criterion for 10% nutrient reduction was within the 5–20% range, with only a 13% probability of complying with the EPA guideline. Reduction by 25 and 50% of the original TP input concentrations would result in 97 and 100% probability to comply with the 10% guideline (Fig. 5). In Hiwassee Reservoir, the exceedance frequency for the 25 μg TP L^{-1} criterion from a 10% nutrient reduction was within the 5–15% range, and thus it already has a high (71%) probability of complying with the 10% EPA guideline.

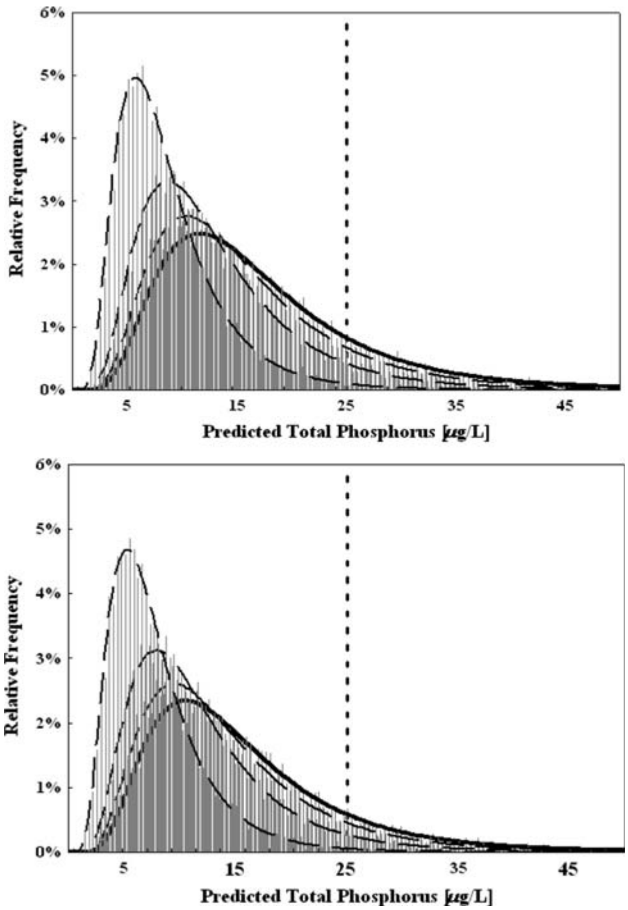


Fig. 3 Distributions of predicted TP_{lake} concentrations for Lake Dudley and Hiwassee Reservoir, respectively. The predicted TP_{lake} concentration distributions from right to left correspond to the original (*solid line*), and 10, 25, 50% (*long dashed lines*) reduction of the TP_{in} concentrations. *Vertical dashed line* corresponds to the 25 μg TP L^{-1} numerical criterion

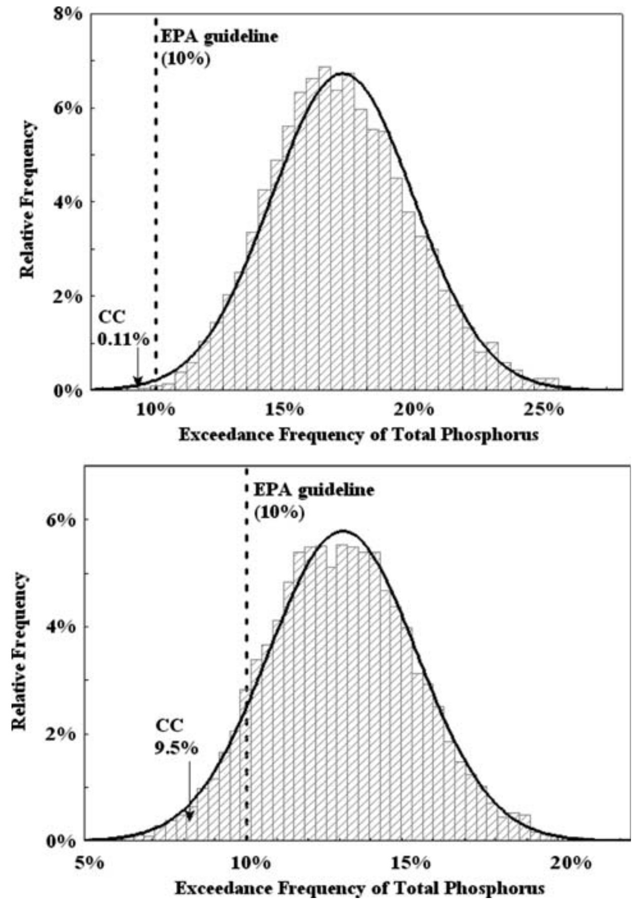


Fig. 4 Probability density for the exceedance frequency of the 25 μg TP L^{-1} numerical criterion for Lake Dudley and Hiwassee Reservoir, respectively. The *dashed line* delineates the area where the exceedance frequency of the numerical criterion is below the 10% EPA guideline and is termed the confidence of compliance (CC)

Reduction by 25 and 50% of the original input concentrations would both result in a 100% probability complying with the EPA guideline (Fig. 5).

Discussion

Empirical models have played an important role in the development of our understanding of the general principles that deal with lake ecosystems (Peters 1986). While their simple conceptual underpinning does not necessarily allow for insights into the complex interplay among the physical, chemical, and biological processes that underlie cultural eutrophication, empirical cross-system relationships in limnology have elucidated large-scale patterns that guide management decisions (Ahlgren et al. 1988). The typical approach in the majority of these cross-sectional studies is to incorporate collateral information or to pool information from different systems and generalize relative to a global trend. These empirical models are then used to make predictions for a single lake at different points in time under the ‘‘debatable’’ assumption that the large-scale behavior described in the model is also representative of within lake dynamics. On top of that, most of these models are missing key regulatory factors (e.g., internal phosphorus loading, top–down control, seasonal patterns in TP losses to the sediments) of the in-lake total phosphorus variability which cast doubt on their ability to effectively support predictions on individual systems (Sarnelle 1999; Søndergaard et al. 2001; Brett and Benjamin 2008). Our objective herein was to revisit the foundation of the existing phosphorus-retention and nutrient-loading models in order to demonstrate how the hierarchical Bayes can be used to balance between site-specific and globally-common model structures. Our hypothesis was that both the plausibility of information pooling and the (possible) evidence of superior performance will affirm the use of hierarchical modeling in the context of water quality management.

Performance of the eutrophication models

Although evidence from the limnological literature suggests that the flow-weighted input phosphorus

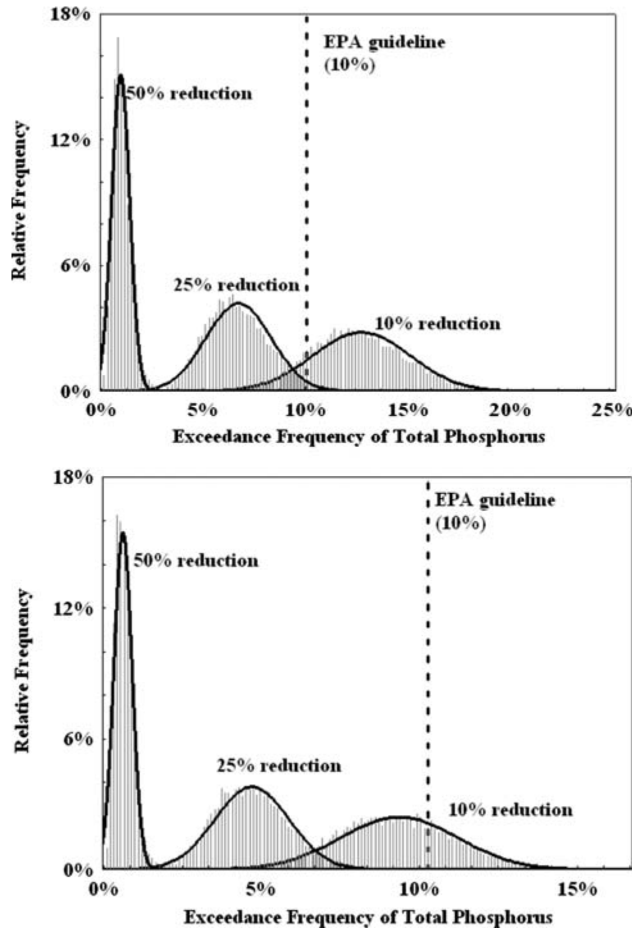


Fig. 5 Probability density for the exceedance frequency of the 25 μg TP L^{-1} numerical criterion with 10, 25, and 50% TP_{in} reduction for Lake Dudley and Hiwassee Reservoir, respectively

concentration is the single best predictor of in-lake TP concentrations (Schindler et al. 1978; Yeasted and Morel 1978; Reckhow 1988), Brett and Benjamin (2008) showed that the $TP_{lake} = f(TP_{in}, k_2)$ model only explains 71% of the overall variability in the log-transformed lake TP concentrations. These results are echoed in our non-hierarchical DIC comparisons, where the corresponding hypothesis (H2) was ranked as the third worst-performing model. On the other hand, the hierarchical configuration of the same model based on the distinction between shallow and deep lakes ($\bar{z} = 10.3$ m) significantly improved its performance ($r^2 = 0.83$) and ended up being the third best model with a hydraulic retention time partitioning at $\tau_w = 2.6$ years. While this result renders support to the implicit assumption of this model that the TP loss rate is approximately proportional to the inverse of the lake hydraulic retention time, it also shows that the slope (k_2) of this relationship is not constant throughout the range examined in this dataset. Namely, the steepness of the $\sigma - 1/\tau_w$ relationship significantly increases in lakes with longer retention times (i.e., flushing rates lower than 38.5% year^{-1}) compared to shorter hydraulic retention time lakes. It is also interesting to note, however, that the empirical relationship H1 that postulates two constant TP sedimentation rates below and above the hydraulic retention time breakpoint (rather than the piecewise linear relationship of the second model) resulted in almost similar performance. The higher sedimentation rates (1.01 ± 0.12 year^{-1}) in lakes with short hydraulic retention times over the rates calculated for longer retention time lakes (0.33 ± 0.3 year^{-1}), along with the k_2 values derived from the H2 model, reiterate the well-documented positive relationship between TP loss and lake flushing rates (Ahlgren et al. 1988). This counter-intuitive relationship was attributed to the fact that the former lakes usually receive relatively greater inputs of allochthonous, mineral-bound (and thus more susceptible to settling) particulate phosphorus than do the latter ones (Schindler et al. 1978; Canfield and Bachmann 1981; Brett and Benjamin 2008).

The hypothesis H4 that the rate coefficient for TP loss from a lake can be expressed as $k_4\tau_{w4}^{x_4-1}$ outperformed the other models in terms of fitting the measured in-lake total phosphorus concentrations (DIC = 469.7 and $r^2 = 0.89$). More importantly, none of the hierarchical models examined led to an improved performance, whereas the group-specific parameters did not differ from the global posteriors or the values obtained from the non-hierarchical approach. Hence, our results verify the conclusions of other studies in that the TP sedimentation rate is best approximated as being proportional to the inverse square root of τ_w, i.e., $\sigma \approx \tau_w^{-0.53}$(see also Fig. 2b in Ahlgren et al. 1988). Similarly to Brett and Benjamin's (2008) assertions, we also found that the same relationship in lakes with shorter hydraulic retention times is better represented by an exponent of 0.37, while the coefficient k_4 that relates the two variables is almost constant throughout the range of hydraulic retention times examined, i.e., $k_{41} \approx k_{42}$. On the other hand, viewing phosphorus sedimentation as a function of the external load as well as the lake phosphorus content (Prairie 1988), hypothesis H5 was the second best performing model with regards to the DIC and r^2 values; especially under the two hierarchical approaches framed upon the mean depth and the hydraulic retention-time classifications. Similar to the other models, the H5 model predicts that shallow and/or shorter hydraulic retention time lakes have higher net sedimentation rates for in-lake phosphorus ($b_1 > 0.45$) compared to deeper lakes or lakes characterized by lower flushing rates ($b_2 < 0.07$). Yet, the group-specific values for the a coefficient also suggest that the fraction of the sedimenting inflowing load is lower in the former lakes ($a_1 > 0.75$) relative to the latter ones ($a_2 < 0.50$). These results offer another perspective in that they emphasize the regulatory role of the in-lake processes rather than the amount of the inflowing allochthonous material to explain the higher net phosphorus sedimentation rates in shallow and/or high-flushing-rate lakes.

Despite the overwhelming support for the particle settling velocity hypothesis (H3) in the limnological literature (Chapra 1975; Dillon and Kirchner 1975; Vollenweider 1976; Ostrofsky 1978; Higgins and Kim 1981; Nürnberg 1984, 1998; Dillon and Molot 1996), our analysis shows that this empirical model is not supported by the present dataset; in fact, the underlying conceptualization did not provide better fit than the much simpler H1 model and only outperforms the two models (H6, H7) that relate TP_{lake} to the TP areal loading. However, we also found that the hierarchical model configuration based on the partitioning between shallow and deep lakes significantly improved the H3 model fit, i.e., $r^2 = 0.87$ relative to the $r^2 = 0.77$ reported in the Brett and Benjamin (2008) study (see their Fig. 4). The same hierarchical approach results in very distinct group-specific posteriors for the apparent settling velocity, i.e., $v_1 = 3.20 \pm 0.38$ m year^{-1} for the shallow and $v_2 = 10.09 \pm 1.04$ m year^{-1} for the deep lakes, respectively. These parameter values are plausible because deep waters are generally less turbulent than surface waters which increases particle aggregation and consequently the effective settling velocity of particulate matter (Malmaeus and Håkanson 2004). The posterior median of v for the deep lakes is very close to the settling velocity value presented by Vollenweider (1975) and falls within the 8–15 m year^{-1} range typically used in other empirical models (e.g., Chapra 1975; Larsen and Mercier 1976; Dillon and Molot 1996). Nonetheless, the particle settling velocity ranges derived herein are still significantly lower (one or two orders of magnitude) than the values reported for phytoplankton cells and detritus (Burns and Rosa 1980; Sommer 1991; Chapra 1997). Brett and Benjamin (2008) presented similar results that were interpreted as additional evidence that the available data do not support the widespread acceptance of the constant particle-settling velocity model. Finally, the two hypotheses that lake TP concentrations are proportional to the areal TP loading or to the ratio between areal nutrient loading and mean lake depth performed quite poorly, although the hierarchical structures based on the lake mean depth (H6) or the hydraulic retention time (H7) classification have notably improved the model fit.

Model-prediction error and suggestions for progress

Striving to improve the existing eutrophication models, Ahlgren et al. (1988) stated that "it is unlikely that these types of models can be further improved only by inclusion of more data in the databases. Instead, it is more likely that careful analyses of homogeneous subsets of data may give models with better predictive value." Indeed, our analysis demonstrated that the hierarchical framework led to an improvement in the performance of six out of the seven hypothesized relationships tested to predict in-lake TP concentrations. While further improvement of the predictive ability of the phosphorus-retention/nutrient-loading models will likely arise from the consideration of lake-specific parameters along with the delineation of more homogeneous groups, founded upon more realistic hierarchical structures (such as the one presented in Fig. 1), the predicted-to-observed TP ratio distributions show that the hierarchical approach still results in considerable predictive uncertainty in individual waterbodies (Fig. 6). Likewise, Malve and Qian's (2006) hierarchical model did not mitigate this problem and, in fact, there were cases in which the poor model fit resulted in negative lake-specific r^2 values (see their Fig. 2). Brett and Benjamin (2008) also showed that these empirical relationships originally fit to TP lake data cannot reproduce observed TP retention values and pinpointed the conditions (e.g., lakes with low areal hydraulic loading rate or long hydraulic retention time) under which they are particularly prone to large prediction errors. Evidently, issues related to the assumptions made for the derivation of these models, and not only the structure used to accommodate the intra- and intersystem variability, are equally important for obtaining robust predictions in individual lakes.

Several studies suggest some of the key underlying model assumptions (e.g., complete mixing, steady state, first-order losses) should be revisited and other unaccounted factors affecting the efficiency of P trapping need to be explicitly considered in these models. These might include the amount of aluminum, iron, and calcium in lake sediments; dominance of the zooplankton community by efficient grazers like *Daphnia*, which can increase the downward flux of phosphorus-containing particulate matter incorporated in fecal pellets; the form in which phosphorus is supplied (i.e., inorganic or organic, dissolved or particulate), which determines if it will be readily incorporated into the biologic cycle of lakes or settle to the sediments; sediment resuspension in shallow, wind-exposed lakes; and bioturbation of material likely to have passed through benthic animals or fish, which creates a "gluing" of the particles into larger flocs (Reynolds and Davies 2001; Brett et al. 2005a, 2005b; Carpenter 2005; Jensen et al. 2006; Brett and Benjamin 2008). Moreover, Nürnberg and LaZerte (2004) indicated that the ratio of watershed-to-lake area can be an important predictor in lakes with complex morphology, while further improvements may be obtained by explicitly considering surrogate variables of soil geochemistry (humic and fulvic acids) such as DOC or lake water color (Nürnberg and Shaw 1998). There is also overwhelming evidence that lake geochemistry influences the retention and settling, e.g., authigenic coprecipitation of P with calcium in calcium-rich systems (Nürnberg 1998).

Eutrophication may also be accentuated due to internal recycling from a large pool of sediment phosphorus, and P sediment release rates have been included in a number of empirical models. For example, Reckhow (1977) differentiated model performance by whether stratified lakes had an oxic or an anoxic hypolimnion. Nürnberg (1984, 1998) showed that Vollenweider-type models only work when the internal load was explicitly added. Nürnberg and LaZerte (2004) explicitly modeled

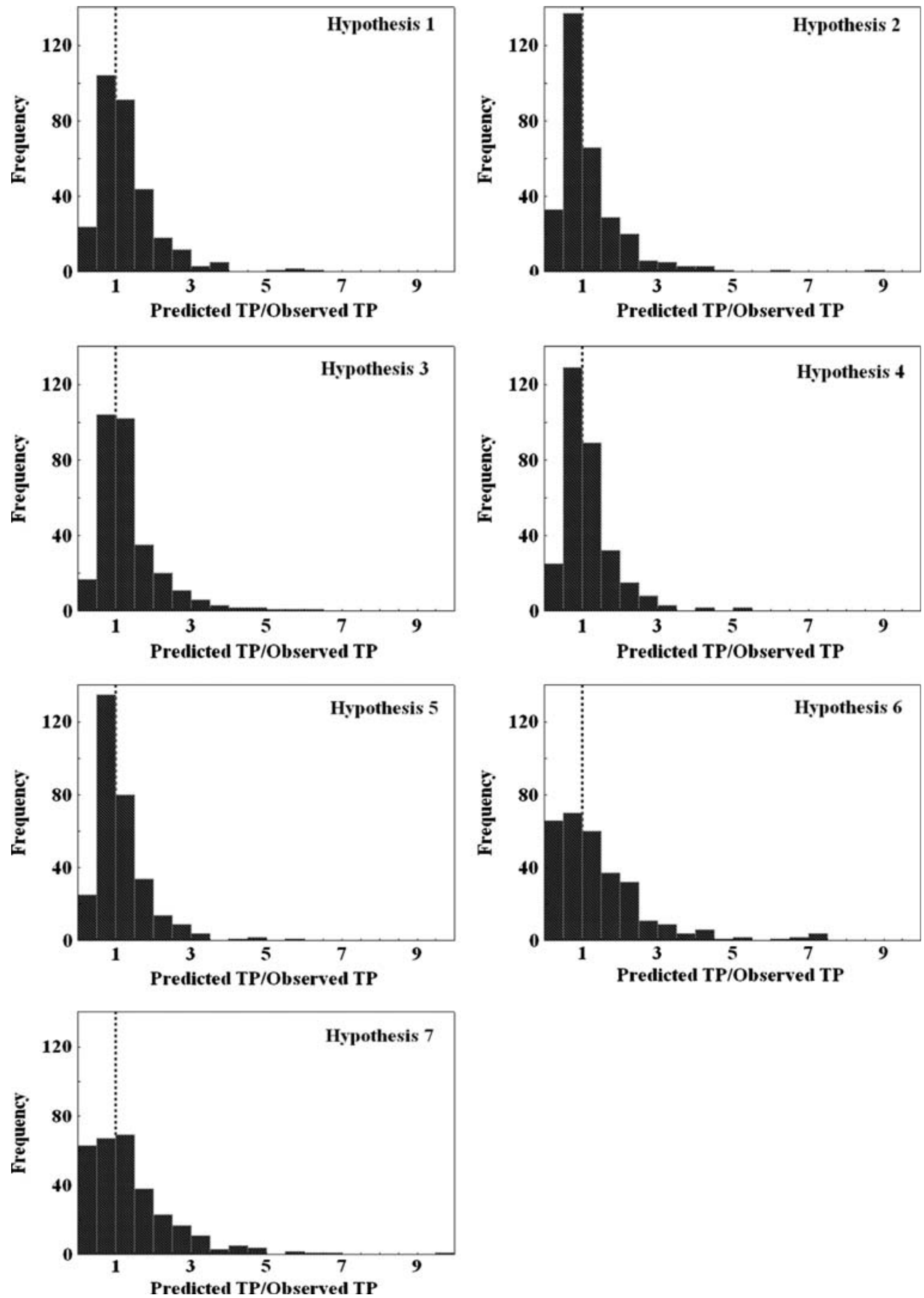


Fig. 6 The distributions of the predicted-to-observed TP ratio for each lake and the best-performing model for each hypothesis assessed (see Table 3). The *hatched vertical line* represents the ideal 1:1 relation

internal P load as the product of sediment release rates and anoxic factors, which were then included in a steady-state phosphorus mass-balance model. The steady-state assumption, however, is inadequate to describe the transient phase following nutrient-loading reduction. Furthermore, phosphorus release from the sediments may prevail for decades; thus, the classical empirical models tend to considerably underestimate in-lake TP concentrations during the recovery phase of eutrophication (Nürnberg 1998; Søndergaard et al. 2003). To address this problem, Jensen et al. (2006) introduced a simple empirical model to describe the early recovery phase by relating the seasonal variation of in-lake TP concentrations to external loading, accumulated phosphorus in the sediment, water temperature, and the hydraulic retention time. These improvements may enable the use of simple models for setting water-quality objectives, while accounting for high internal loading and thus unanticipated delays in system recovery.

The Bayesian nature of the proposed hierarchical framework has several advantages for environmental management. For the purposes of probabilistic risk assessment, the Bayesian approach generates a posterior predicted distribution that represents the current estimate of in-lake TP concentrations, taking into account both the uncertainty for the parameters and the uncertainty that remains when the parameters are known (Gelman et al. 1995). Therefore, estimates of the uncertainty of Bayesian-model predictions are more realistic (usually larger) than those based on the classical procedures and the target nutrient loads can be set by explicitly acknowledging an inevitable risk of non-attainment. For example, our analysis illustrated how this methodological framework can be used for assessing the exceedance frequency and confidence of compliance of different water quality standards (Figs. 3–5). The Bayesian approach also provides a convenient means to make decisions about nutrient-load reductions that reflect different socioeconomic values and environmental concerns, i.e., management objectives can be evaluated by integrating the probability of use attainment for a given water quality goal with utility functions; the management scheme associated with the highest expected utility might then be chosen (Dorazio and Johnson 2003). Finally, other benefits for environmental management include alignment with the policy practice of adaptive management implementation (Malve and Qian 2006), and the optimization of monitoring programs using value of information concepts from decision theory (Arhonditsis et al. 2007; Zhang and Arhonditsis 2008).

In conclusion, we presented a Bayesian hierarchical framework that relaxes the basic assumption of empirical models fitted to cross-sectional data sets, i.e., the systems are assumed to be identical in behavior, and therefore the models have a single globally common set of parameters. The hierarchical structure enables the estimation of group and/or system-specific parameters that explicitly consider the role of significant sources of variability (e.g., morphometry, mixing regime, geographical location, land-use patterns, and trophic status). Our analysis showed that the hierarchical framework improves the performance of the phosphorus-retention/nutrient-loading models, although predictive statements for individual lakes still have large error. A hypothesis based on the assumption that the first-order TP loss rate can be expressed as $\sigma = k_4 \tau_{w4}^{x-1}$ predicted the observed in-lake total phosphorus concentrations better than any of the other relationships examined. In particular, the TP sedimentation rate is best approximated as being proportional to the inverse square root of τ_w, i.e., $\sigma \approx 1.12\tau_w^{-0.53}$, although some variability exists with regards to the values of the power x_4 depending on the lake hydraulic retention time. The particle-settling velocity hypothesis that estimates σ as the ratio of a nominal particle settling velocity to the lake mean depth was not strongly supported by the present dataset. However, the performance of this empirical model considerably improves when different settling rates are assigned to shallow and deep lakes. Future improvements should also focus on building more realism into the models by explicitly quantifying recycling fluxes and transient dynamics of lakes following P-load reductions. The proposed methodological framework has several advantages that are useful for assessing water-quality conditions and will facilitate the policy-making process.

Acknowledgments Funding for this study was provided by the UTSC Research Fellowships (Master of Environmental Science Program, Centre for Environment and University of Toronto at Scarborough) and the National Sciences and Engineering Research Council of Canada (NSERC, Discovery Grants). Gertrud Nurnberg and one anonymous reviewer offered excellent comments on an earlier version of the manuscript.

References

Ahlgren I, Frisk T, Kamp-Nielsen L (1988) Empirical and theoretical models of phosphorus loading, retention, and concentration vs. lake trophic state. Hydrobiologia 170:285–303

Amrhein JF, Stow CA, Wible C (1999) Whole-fish versus filet polychlorinated-biphenyl concentration: an analysis using classification and regression tree models. Environ Toxicol Chem 18:1817–1823

Arhonditsis GB, Qian SS, Stow CA, Lamon EC, Reckhow KH (2007) Eutrophication risk assessment using Bayesian calibration of process-based models: application to a mesotrophic lake. Ecol Modell 208:215–229

Borsuk ME, Higdon D, Stow CA, Reckhow KH (2001) A Bayesian hierarchical model to predict benthic oxygen demand from organic matter loading in estuaries and coastal zones. Ecol Modell 143:165–181

Borsuk ME, Stow CA, Reckhow KH (2002) Predicting the frequency of water quality standard violations: a probabilistic approach for TMDL development. Environ Sci Technol 36:2109–2115

Breiman L, Friedman JH, Olshen RA, Stone CJ (1984) Classification and regression trees. Wadsworth International Group, Belmont

Brett MT, Benjamin MM (2008) A review and reassessment of lake phosphorus retention and the nutrient loading concept. Freshw Biol 53:194–211

Brett MT, Arhonditsis GB, Mueller SE, Hartley DM, Frodge JD, Funke DE (2005a) Non-point source nutrient impacts on stream nutrient and sediment concentrations along a forest to urban gradient. Environ Manage 35:330–342

Brett MT, Mueller SE, Arhonditsis GB (2005b) A daily time series analysis of stream water phosphorus transport along an urban to forest gradient in the Seattle area. Environ Manage 35:56–71

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

Burns NM, Rosa F (1980) In situ measurement of the settling velocity of organic-carbon particles and 10 species of phytoplankton. Limnol Oceanogr 25:855–864

Canfield DE Jr, Bachmann RW (1981) Prediction of total phosphorus concentrations, chlorophyll *a*, and Secchi depths in natural and artificial lakes. Can J Fish Aquat Sci 38:414–423

Carpenter SR (2005) Eutrophication of aquatic ecosystems: bistability and soil phosphorus. Proc Natl Acad Sci USA 102:10002–10005

Chapra SC (1975) Comment on 'An empirical method of estimating the retention of phosphorus in lakes' by WB Kirchner and PJ Dillon. Water Resour Res 11:1033–1034

Chapra SC (1997) Surface water-quality modeling. McGraw-Hill, New York

Chapra SC, Reckhow KH (1979) Expressing the phosphorus loading concept in probabilistic terms. J Fish Res Board Can 36:225–229

Clark JS (2005) Why environmental scientists are becoming Bayesians. Ecol Lett 8:2–14
Clark LA, Pregibon D (1992) Tree based models. In: Chambers JM, Hastie TJ (eds) Statistical models in S. Wadsworth and Brooks/Cole Advanced Books and Software, Pacific Grove, pp 377–420
Clark JS, Dietze M, Chakraborty S, Agarwal PK, Ibanez I, LaDeau S, Wolosin M (2007) Resolving the biodiversity paradox. Ecol Lett 10:647–659
De'ath G, Fabricius KE (2000) Classification and regression trees: a powerful yet simple technique for ecological data analysis. Ecology 81:3178–3192
Dillon PJ, Kirchner WB (1975) Reply to Chapra's comment. Water Resour Res 11:1035–1036
Dillon PJ, Molot LA (1996) Long-term phosphorus budgets and an examination of a steady-state mass balance model for central Ontario lakes. Water Res 30:2273–2280
Dorazio RM, Johnson FA (2003) Bayesian inference and decision theory—a framework for decision making in natural resource management. Ecol Appl 13:556–563
Ellison AM (1996) An introduction to Bayesian inference for ecological research and environmental decision-making. Ecol Appl 6:1036–1046
Fricker H (1980) OECD eutrophication programme regional project Alpine lakes. Swiss Federal Board for Environmental Protection & OECD, Dübendorf
Gelman A, Hill J (2007) Data analysis using regression and multilevel/hierarchical models, second printing. Cambridge University Press, New York
Gelman A, Carlin JB, Stern HS, Rubin DB (1995) Bayesian data analysis. Chapman & Hall, New York, p 518
Gilks W, Roberts GO, Sahu SK (1998) Adaptive Markov chain Monte Carlo through regeneration. J Am Stat Assoc 93:1045–1054
Higgins JM, Kim BR (1981) Phosphorus retention models for Tennessee Valley Authority reservoirs. Water Resour Res 17:571–576
Janus LL, Vollenweider RA (1981) Summary report, the OECD cooperative programme on eutrophication report, Canadian contribution. Cananda Centre for Inland Waters, Burlington
Jensen JP, Pedersen AR, Jeppesen E, Søndergaard M (2006) An empirical model describing the seasonal dynamics of phosphorus in 16 shallow eutrophic lakes after external loading reduction. Limnol Oceanogr 51:791–800
Jones JR, Bachman RW (1976) Prediction of phosphorus and chlorophyll levels in lakes. J Water Pollut Control Fed 48:2176–2182
Judge GG, Griffiths WE, Carter Hill R, Lütkepohl H, Lee TC (1985) The theory and practice of econometrics, 2nd edn. Wiley, New York
Kirchner WB, Dillon PJ (1975) Empirical method of estimating retention of phosphorus in lakes. Water Resour Res 11:182–183
Lamon EC, Stow CA (1999) Sources of variability in microcontaminant data for Lake Michigan salmonids: statistical models and implications for trend detection. Can J Fish Aquat Sci 56(Suppl 1):71–85
Lamon EC, Stow CA (2004) Bayesian methods for regional-scale eutrophication models. Water Res 38:2764–2774
Larsen DP, Mercier HT (1976) Phosphorus retention capacity of lakes. J Fish Res Board Can 33:1742–1750
Lunn DJ, Thomas A, Best N, Spiegelhalter D (2000) WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility. Stat Comput 10:325–337
Magnuson JJ, Tonn WN, Banjeree A, Toivonnen J, Sanchez O, Rask M (1998) Isolation vs. extinction in the assembly of fishes in small northern lakes. Ecology 79:2941–2956
Malmaeus JM, Håkanson L (2004) Development of a lake eutrophication model. Ecol Modell 171:35–63
Malve O, Qian SS (2006) Estimating nutrients and chlorophyll a relationships in Finnish lakes. Environ Sci Technol 40:7848–7853
Michielsens C, McAllister M (2004) A Bayesian hierarchical analysis of stock-recruit data: quantifying structural and parameter uncertainties. Can J Fish Aquat Sci 61:1032–1047
Nerini D, Dubrec JP, Mante C (2000) Analysis of oxygen rate time series in a strongly polluted lagoon using a regression tree method. Ecol Modell 133:95–105
Nürnberg GK (1984) The prediction of internal phosphorus load in lakes with anoxic hypolimnia. Limnol Oceanogr 29:111–124
Nürnberg GK (1996) Trophic state of clear and colored, soft- and hard- water lakes with special consideration of nutrients, anoxia, phytoplankton and fish. Lake Reserv Manage 12:432–447
Nürnberg GK (1998) Prediction of annual and seasonal phosphorus concentration in stratified and polymictic lakes. Limnol Oceanogr 43:1544–1552
Nürnberg GK, LaZerte BD (2004) Modeling the effect of development on internal phosphorus load in nutrient-poor lakes. Water Resour Res 40:W01105. doi:10.1029/2003WR002410
Nürnberg GK, Shaw M (1998) Productivity of clear and humic lakes: nutrients, phytoplankton, bacteria. Hydrobiologia 382:97–112
Ostrofsky ML (1978) Modification of phosphorus retention models for use with lakes with low areal water loading. J Fish Res Board Can 35:1532–1536
Peters RH (1986) The role of prediction in limnology. Limnol Oceanogr 31:1143–1159
Prairie YT (1988) A test of the sedimentation assumptions of phosphorus input–output models. Arch Hydrobiol 111:321–327
Prairie YT (1989) Statistical models for the estimation of net phosphorus sedimentation in lakes. Aquat Sci 51:192–210
Prairie YT, Marshall CT (1995) On the use of structured time-series to detect and test hypotheses about within-lakes relationships. Can J Fish Aquat Sci 52:799–803
Prairie YT, Peters RH, Bird DF (1995) Natural variability and the estimation of empirical relationships—a reassessment of regression methods. Can J Fish Aquat Sci 52:788–798
Qian SS, Anderson CW (1999) Exploring factors controlling variability of pesticide concentrations in the Willamette River Basin using tree-based models. Environ Sci Technol 33:3332–3340
Rast W, Lee GF (1978) Summary analysis of the North American (U.S. portion) OECD eutrophication programme: nutrient loading-lake response relationship and trophic state indices. Ecological Research Series, No. EPA-600/3-78-008, U.S. Environmental Protection Agency, Springfield, Virginia, USA
Reckhow KH (1977) Phosphorus models for lake management. PhD Thesis, Harvard University, Cambridge, MA. pp 304
Reckhow KH (1988) Empirical models for trophic state in southeastern U.S. lakes and reservoirs. Water Res Bull 24:723–734
Reckhow KH (1993) A random coefficient model for chlorophyll–nutrient relationships in lakes. Ecol Modell 70:35–50
Reckhow KH, Chapra SC (1983) Engineering approaches for lake management. Vol. 1: data analysis and empirical modeling. Butterworth Publishers, Woburn
Reynolds CS, Davies PS (2001) Sources and bioavailability of phosphorus fractions in freshwaters: a British perspective. Biol Rev 76:27–64
Rivot E, Prévost E, Cuzol A, Baglinière J, Parent E (2008) Hierarchical Bayesian modelling with habitat and time covariates for estimating riverine fish population size by successive removal method. Can J Fish Aquat Sci 65:117–133
Ryding SO (1980) Monitoring of inland waters. OECD eutrophication programme. The Nordic project, Nordforsk, Helsinki
Sarnelle O (1999) Zooplankton effects on vertical particulate flux: testable models and experimental results. Limnol Oceanogr 44:357–370
Schindler DW, Fee EJ, Ruszczynski T (1978) Phosphorus input and its consequences for phytoplankton standing crop and production in the Experimental Lakes Area and in similar lakes. J Fish Res Board Can 35:190–196
Snodgrass WJ, O'Melia CR (1975) Predictive model for phosphorus in lakes. Environ Sci Technol 9:937–944

Sommer U (1991) Phytoplankton: directional succession and forced cycles. In: Remmert H (ed) The mosaic-cycle concept of ecosystems. Springer, Berlin Heidelberg New York, pp 132–146
Søndergaard M, Jensen JP, Jeppesen E (2001) Retention and internal loading of phosphorus in shallow, eutrophic lakes. The Scientific World 1:427–442
Søndergaard M, Stedmon C, Borch NH (2003) Fate of terrestrial dissolved organic carbon in estuaries: aggregation and bioavailability. Ophelia 57:161–176
Spiegelhalter D, Best N, Carlin B, van der Linde A (2002) Bayesian measures of model complexity and fit. J Roy Stat Soc B 64:583–639 with discussion 699, 700
Swamy PAVB (1971) Statistical inference in random coefficient regression models. Springer, Berlin Heidelberg New York
USEPA, 1975. A compendium of lakes and reservoir data collected by the national eutrophication survey in the northeast and northcentral United States. US Environmental Protection Agency working paper No. 474. Corvallis, Oregon, USA
Uttormark PD, Hutchins ML (1978) Input–output models as decision criteria for lake restoration. Technical report, Wis, Water Resources Center 78–03, pp 61
Vollenweider RA (1968) The scientific basis of lake and stream eutrophication, with particular reference to phosphorus and nitrogen as eutrophication factors. Paris: Organization for Economic Co-operation and Development. Technical report: DAS/CSI/68.27. pp 250
Vollenweider RA (1969) Möglichkeiten und grenzen elementarer modelle der stoffbilanz von seen. arch. Hydrobiol 66:1–36
Vollenweider RA (1975) Input–output models with special reference to the phosphorus loading concept in limnology. Schweiz Z Hydrol 37:53–84
Vollenweider RA (1976) Advances in defining critical loading levels for phosphorus in lake eutrophication. Mem Ist Ital Idrobiol 33:58–83
Walker WW Jr. (1977) Some analytical methods applied to lake water quality problems. PhD Thesis, Harvard University, Cambridge, MA
Welch EB (1992) Ecological effects of wastewater: applied limnology and pollutant effects, 2nd edn. Chapman & Hall, New York
Wikle CK (2003a) Hierarchical models in environmental science. Int Stat Rev 71:181–199
Wikle CK (2003b) Hierarchical Bayesian models for predicting the spread of ecological processes. Ecology 84:1382–1394
Wyatt RJ (2002) Estimating riverine fish population size from single- and multiple-pass removal sampling using a hierarchical model. Can J Fish Aquat Sci 59:695–706
Yeasted JG, Morel FMM (1978) Empirical insights into lake response to nutrient loadings, with application to models of phosphorus in lakes. Environ Sci Technol 12:195–201
Zhang W, Arhonditsis GB (2008) Predicting the frequency of water quality standard violations using Bayesian calibration of eutrophication models. J Great Lakes Res 34:698–720
Zhao J, Ramin M, Cheng V, Arhonditsis GB (2008a) Plankton community patterns across a trophic gradient: the role of zooplankton functional groups. Ecol Modell 213:417–436
Zhao J, Ramin M, Cheng V, Arhonditsis GB (2008b) Competition patterns among phytoplankton functional groups: how useful are the complex mathematical models? Acta Oecol 33:324–344

A REVISIT OF LAKE PHOSPHORUS LOADING MODELS USING A BAYESIAN HIERARCHICAL FRAMEWORK

(Electronic Supplementary Material)

Vincent Cheng, George B. Arhonditsis*

Department of Physical & Environmental Sciences, University of Toronto,

Toronto, Ontario, Canada, M1C 1A4

and Michael T. Brett

Department of Civil & Environmental Engineering, University of Washington,

Seattle, Washington, USA 98195

* Corresponding author

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

Table 1-ESM. General limnological characteristics of the lakes included in this analysis.

Source	*n*	*Percentile*	*Volume* *(10^6 m^3)*	*Surface area* *(km^2)*	*Mean depth* *(m)*	*Inflow rate* *(m^{3}s^{-1})*	*Areal hydraulic loading* *(m yr^{-1})*	*Hydraulic retention time* *(yrs)*	*TP_{in}* *(µg L^{-1})*	*TP_{lake}* *(µg L^{-1})*	*TP_{lake}/TP_{in}* *(unitless)*
Overall	305	90th	2487	127	33.6	150	207	6.25	398	213	1.04
Overall	305	75th	347.2	27.0	14.3	20.3	56.5	2.60	194	92.0	0.82
Overall	305	50th	48.5	7.2	6.4	3.6	13.8	0.58	72.7	33.0	0.55
Overall	305	25th	4.7	1.3	3.2	0.5	4.1	0.10	31.0	12.1	0.31
Overall	305	10th	1.1	0.4	1.7	0.1	1.4	0.02	15.6	7.8	0.18
Arithmetic Mean	305		76059	841	16.2	134	96.4	5.80	179	92.8	0.60
Geometric Mean	305		52.4	7.2	7.3	3.6	15.6	0.47	76.6	37.2	0.49
Skewness	305		13.4	9.6	5.0	7.6	10.4	15.3	9.4	4.7	0.78
Kurtosis	305		197	97	33.2	68.8	138	250	104	27.9	0.56
Vollenweider (1969)	8	50th	233	7.9	31.0	1.8	7.4	3.22	122	64.0	0.55
NES (1975)	134	50th	28.9	6.1	4.4	4.9	19.2	0.22	90.1	50.0	0.64
Jones and Bachman (1976)	16	50th	5.2	3.5	2.0	0.2	1.3	1.60	259	86.5	0.34
Rast and Lee (1978)	30	50th	15.2	2.7	6.2	0.6	4.0	2.05	109	50.0	0.41
Fricker (1980)	20	50th	163	7.9	34.2	2.3	17.9	1.54	72.5	23.9	0.35
Rydin (1980)	14	50th	282	24.1	14.7	15.2	17.2	1.08	63.7	25.7	0.50
Higgins and Kim (1981)	18	50th	813	70.7	10.1	350	133	0.08	38.9	28.0	0.79
Janis and Vollenweider (1981)	65	50th	62.2	5.2	10.5	2.0	10.9	0.94	27.7	10.5	0.54

FIGURES LEGENDS

Figure 1-ESM: Marginal posterior distributions of the model error (τ) for the non-hierarchical (dash line) and the corresponding hierarchical models partitioned with mean depth - $Prior_3$ (solid line).

Figure 2-ESM: Marginal posterior distributions of the model error (τ) for the non-hierarchical (dash line) and the corresponding hierarchical models partitioned with hydraulic retention - $Prior_3$ (solid line).

Figure 3-ESM: Sensitivity of the posterior patterns on the specification of the global prior distributions for the parameter k_1 (*H1* model). The priors are normally distributed with 95, 80, 68.2, 50, 34.1, 25% of their values lying within the interval 0.45±0.04 yr^{-1} reported by Brett and Benjamin (2008).

Figure 4-ESM: Observed versus median predicted lake total phosphorus (TP_{lake}) concentrations [μg L^{-1}]. Gray lines correspond to the 2.5 and 97.5% credible intervals. The diagonal line represents a perfect fit between predicted medians and observed values.

Figure 5-ESM: Bayesian hierarchical model based on the hydraulic retention partitioning. Prior (thick black lines) and posterior group-specific (thin black: ≤ τ_w = 2.6 yrs, and thick gray: > τ_w = 2.6 yrs) parameter distributions. Global priors are normally distributed within the confidence intervals reported in the Brett and Benjamin (2008) study ($Prior_3$).

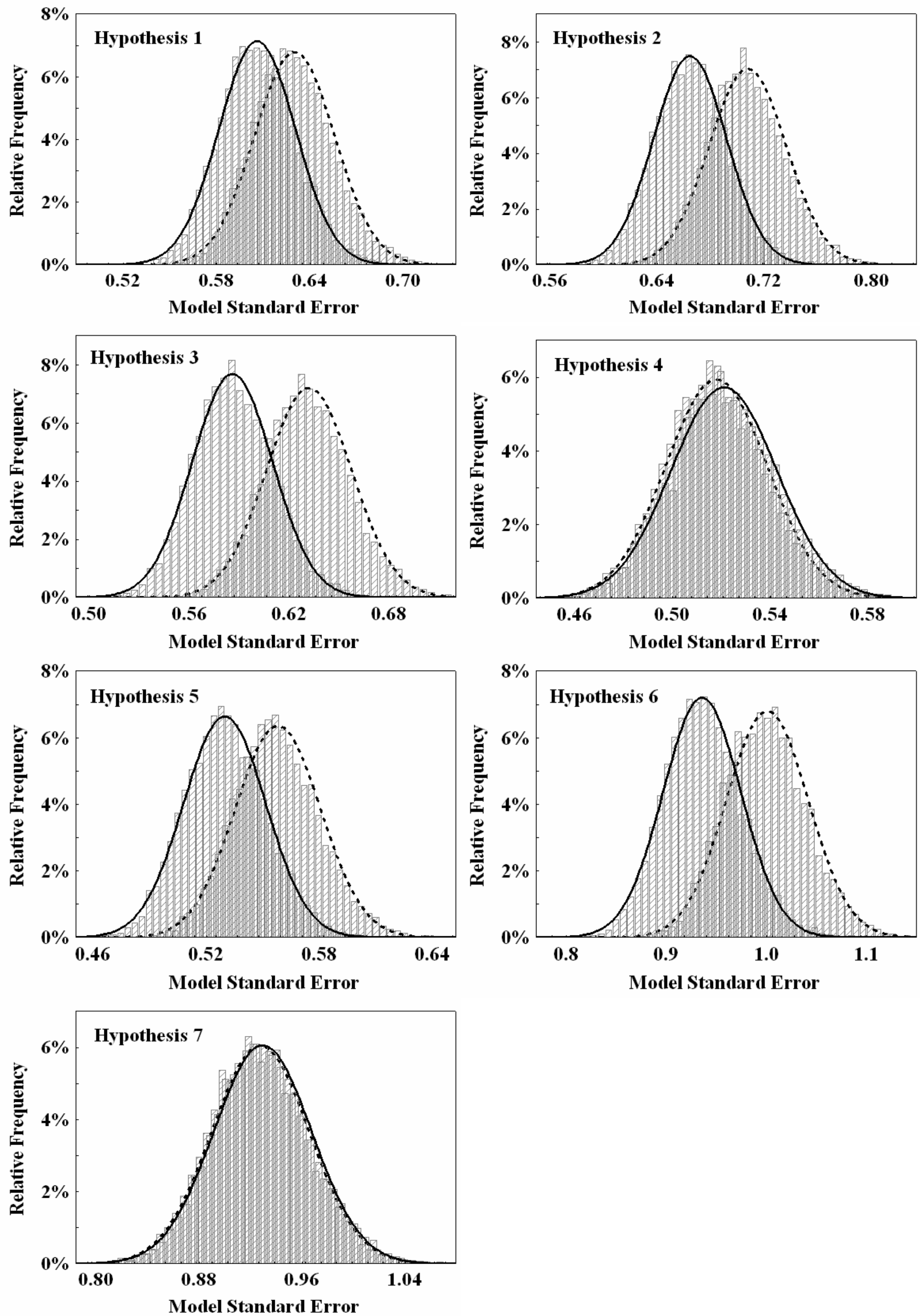
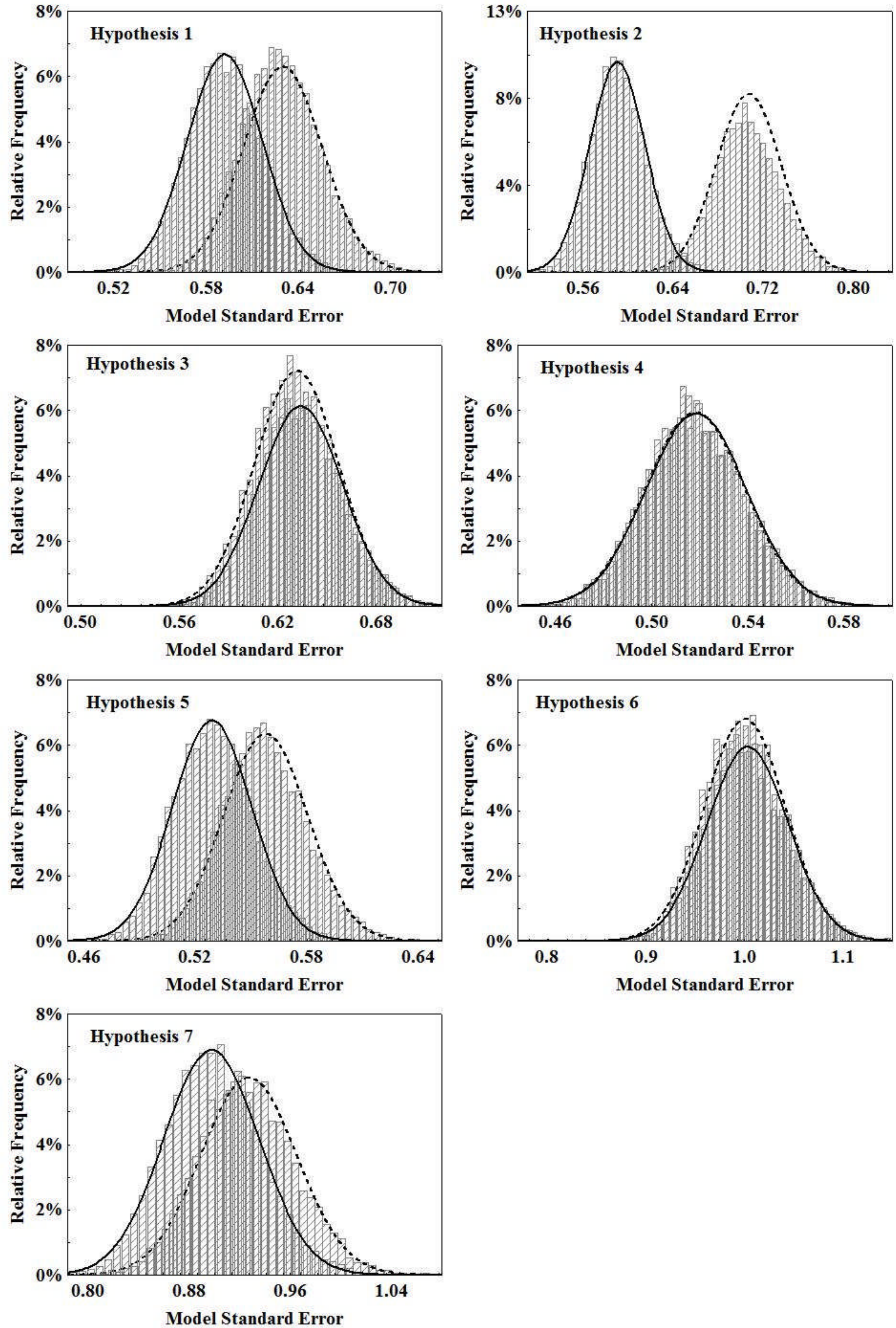

Hypothesis 1
Hypothesis 2
Hypothesis 3
Hypothesis 4
Hypothesis 5
Hypothesis 6
Hypothesis 7
Relative Frequency
Model Standard Error
0%
2%
4%
6%
8%
0.52
0.58
0.64
0.70
0.56
0.64
0.72
0.80
0.50
0.56
0.62
0.68
0.46
0.50
0.54
0.58
0.46
0.52
0.58
0.64
0.8
0.9
1.0
1.1
0.80
0.88
0.96
1.04

Figure 1-ESM



Hypothesis 1
Hypothesis 2
Hypothesis 3
Hypothesis 4
Hypothesis 5
Hypothesis 6
Hypothesis 7
Relative Frequency
Model Standard Error

Figure 2-ESM

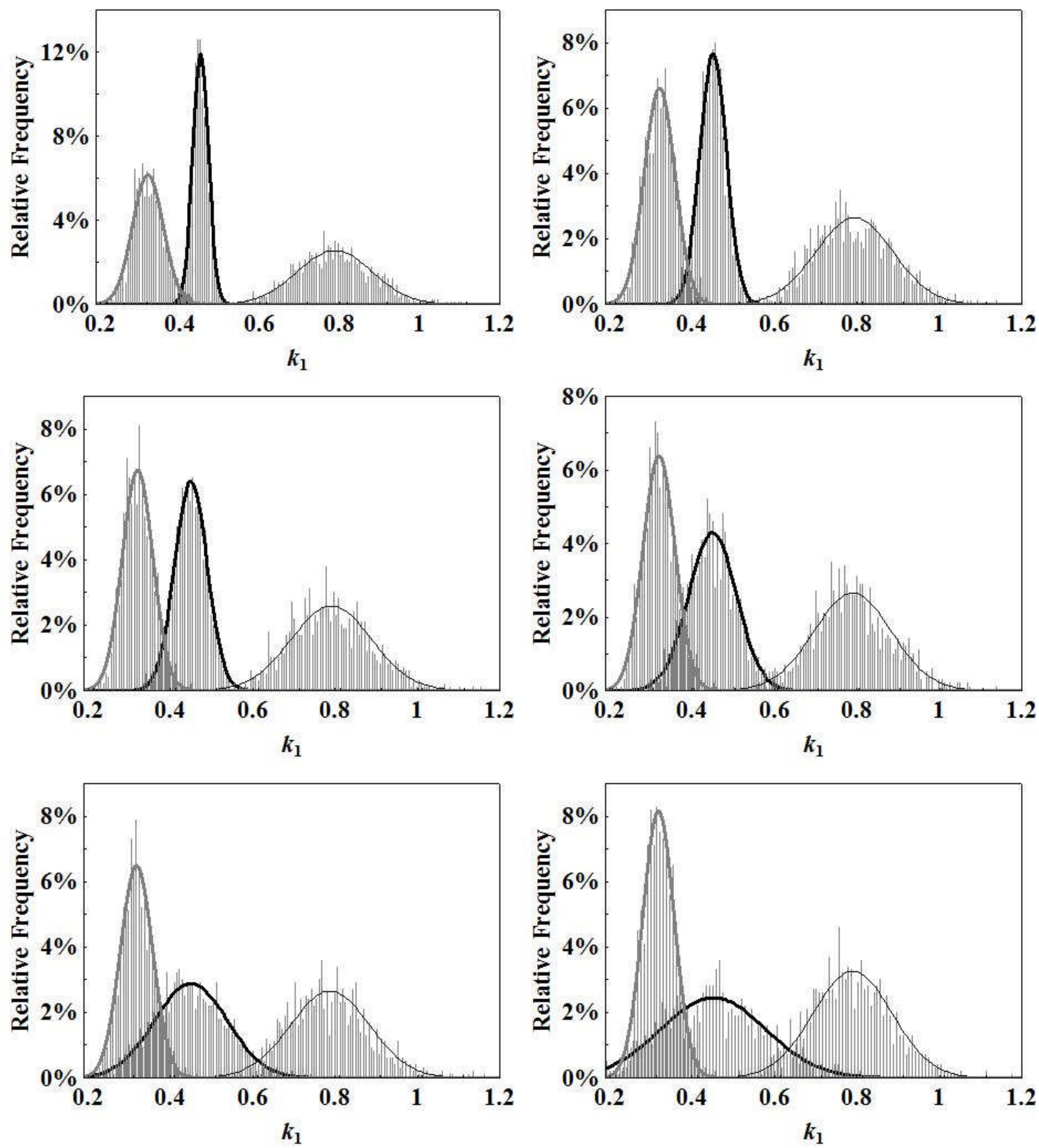


Figure 3-ESM

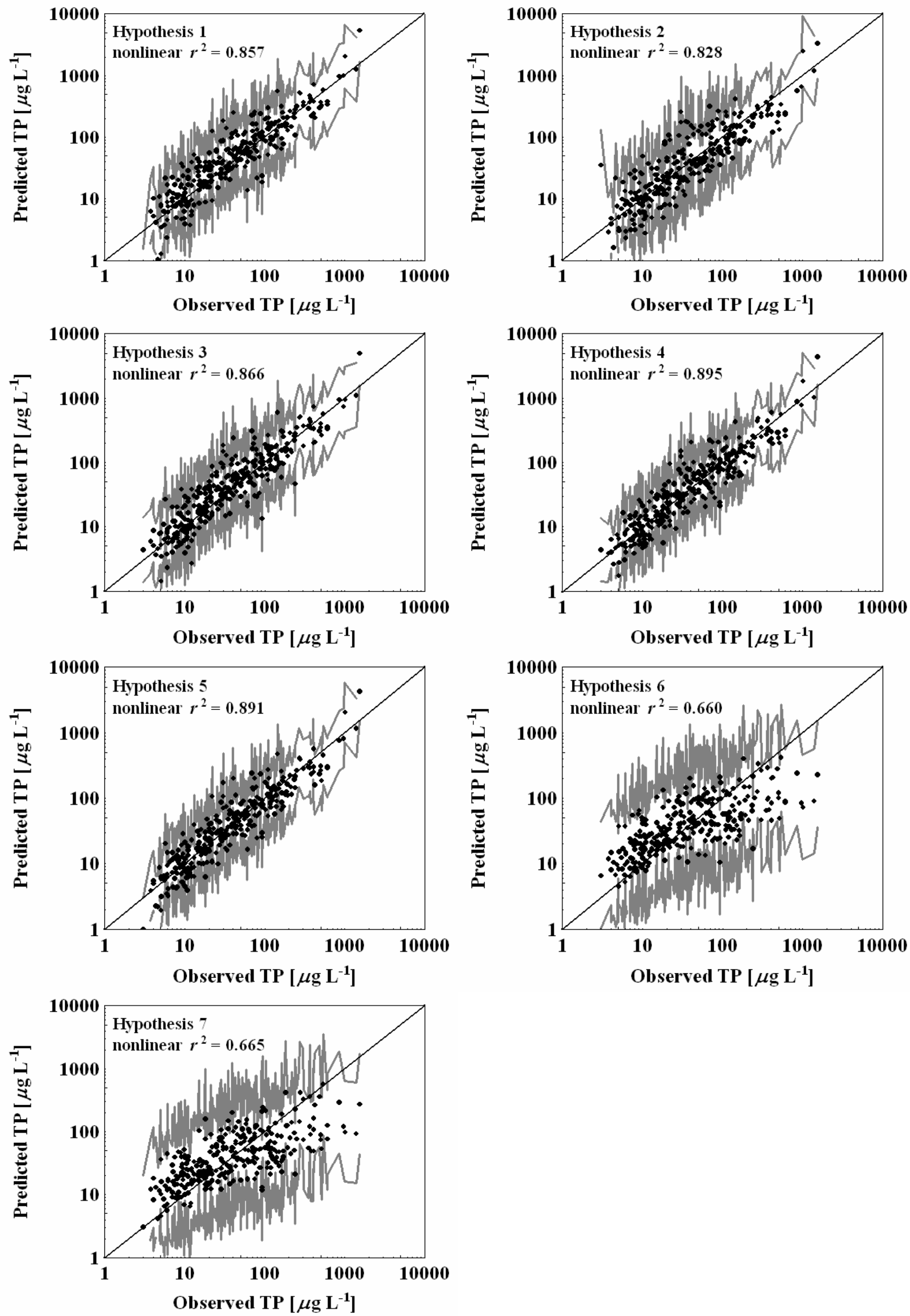
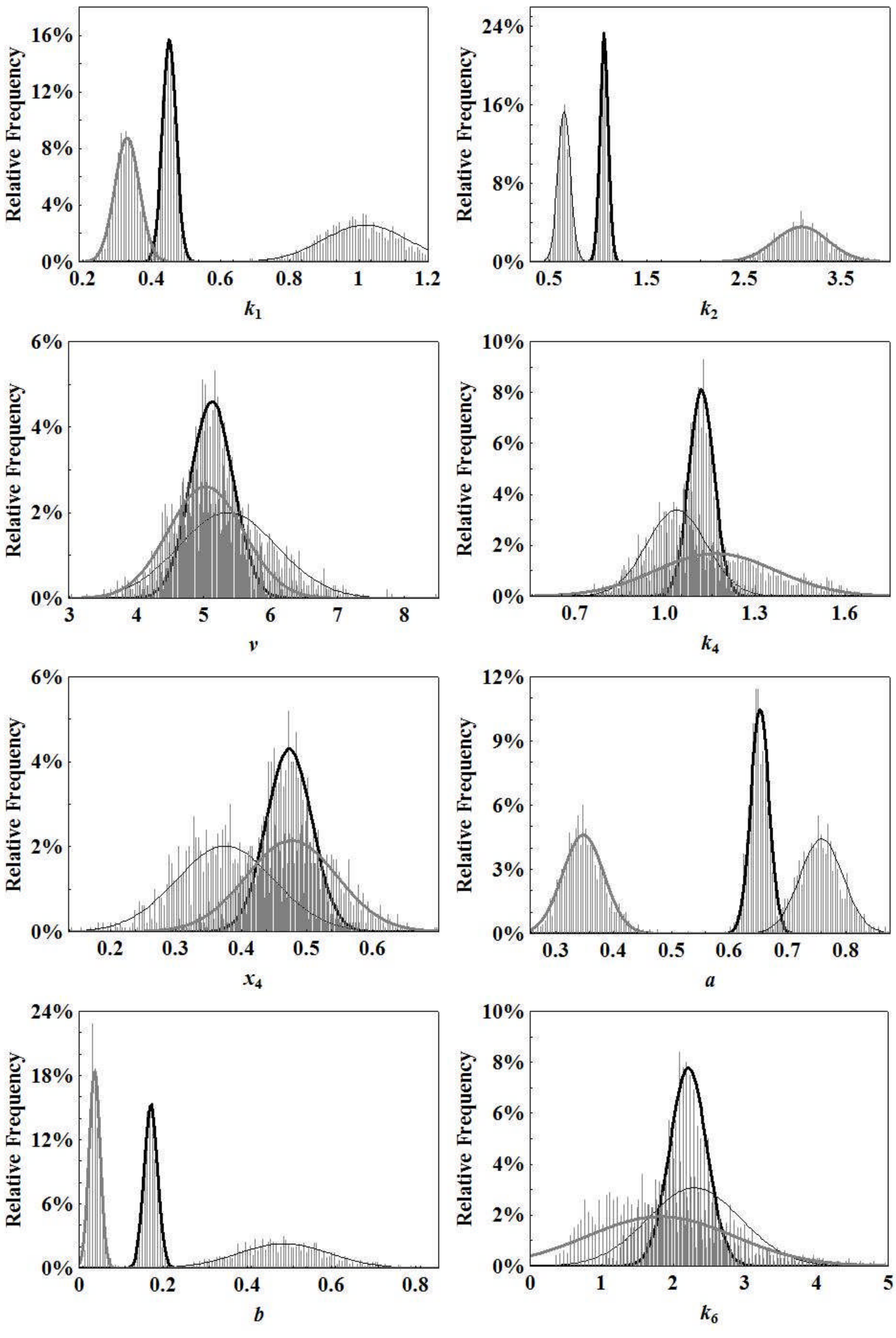

Hypothesis 1
nonlinear $r^2 = 0.857$
Hypothesis 2
nonlinear $r^2 = 0.828$
Hypothesis 3
nonlinear $r^2 = 0.866$
Hypothesis 4
nonlinear $r^2 = 0.895$
Hypothesis 5
nonlinear $r^2 = 0.891$
Hypothesis 6
nonlinear $r^2 = 0.660$
Hypothesis 7
nonlinear $r^2 = 0.665$
Predicted TP [$\mu g\ L^{-1}$]
Observed TP [$\mu g\ L^{-1}$]
1
10
100
1000
10000

Figure 4-ESM


Relative Frequency
16%
12%
8%
4%
0%
0.2 0.4 0.6 0.8 1 1.2
k1
24%
16%
8%
0%
0.5 1.5 2.5 3.5
k2
6%
4%
2%
0%
3 4 5 6 7 8
v
10%
8%
6%
4%
2%
0%
0.7 1.0 1.3 1.6
k4
6%
4%
2%
0%
0.2 0.3 0.4 0.5 0.6
x4
12%
9%
6%
3%
0%
0.3 0.4 0.5 0.6 0.7 0.8
a
24%
18%
12%
6%
0%
0 0.2 0.4 0.6 0.8
b
10%
8%
6%
4%
2%
0%
0 1 2 3 4 5
k6

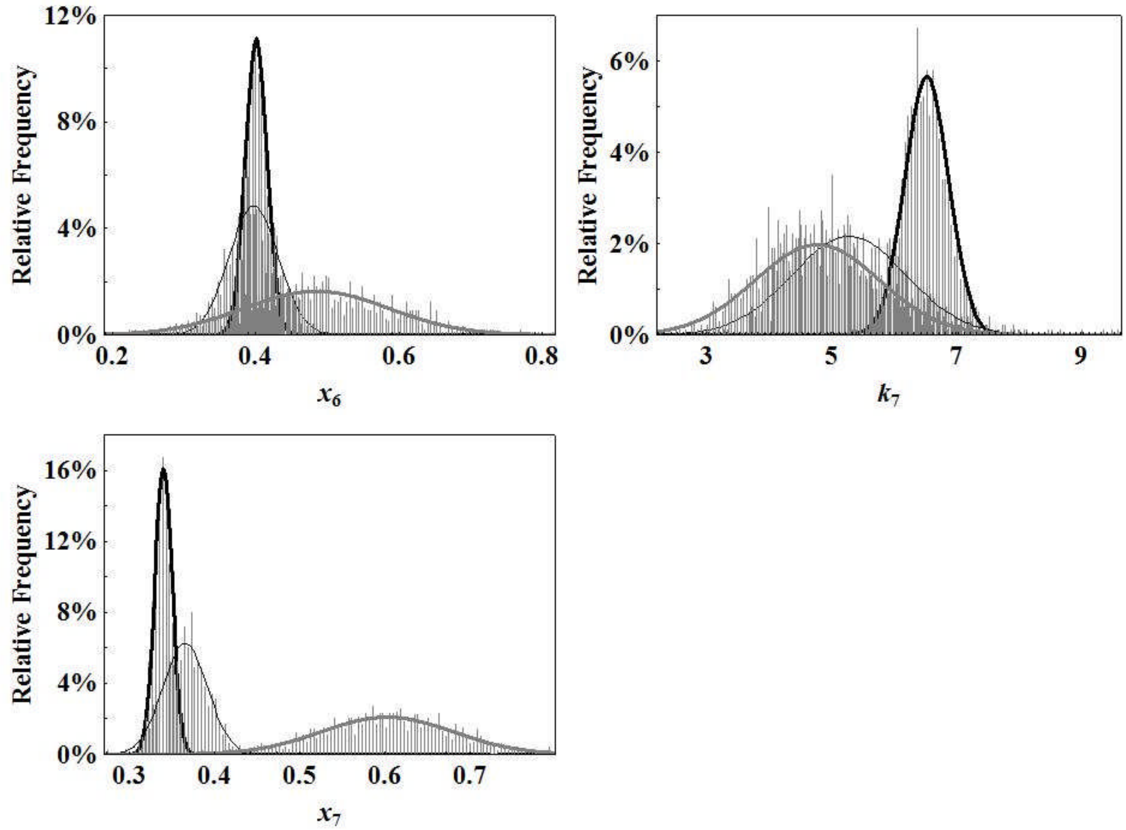

12%
8%
4%
0%
Relative Frequency
0.2
0.4
0.6
0.8
x_6
6%
4%
2%
0%
Relative Frequency
3
5
7
9
k_7
16%
12%
8%
4%
0%
Relative Frequency
0.3
0.4
0.5
0.6
0.7
x_7

Figure 5-ESM