Ecological Modelling 242 (2012) 127–145

Contents lists available at SciVerse ScienceDirect

Ecological Modelling

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

A Bayesian synthesis of predictions from different models for setting water quality criteria

Maryam Ramin[a], Tanya Labencki[b], Duncan Boyd[b], Dennis Trolle[c], George B. Arhonditsis[a,*]

[a] Ecological Modelling Laboratory, Department of Physical & Environmental Sciences, University of Toronto, Toronto, Ontario, Canada M1C 1A4
[b] Great Lakes Unit, Water Monitoring & Reporting Section, Ontario Ministry of the Environment, Environmental Monitoring and Reporting Branch, Toronto, Ontario, Canada M9P 3V6
[c] Department of Freshwater Ecology, National Environmental Research Institute, Aarhus University, PO Box 314, 8600 Silkeborg, Denmark

ARTICLE INFO

Article history:
Received 10 April 2012
Received in revised form 24 May 2012
Accepted 26 May 2012

Keywords:
Process-based modelling
Eutrophication
Bayesian inference
Model averaging
Water quality criteria
Decision making
Sediment diagenesis

ABSTRACT

Skeptical views of the scientific value of modelling argue that there is no true model of an ecological system, but rather several adequate descriptions of different conceptual basis and structure. In this regard, rather than picking the single "best-fit" model to predict future system responses, we can use Bayesian model averaging to synthesize the forecasts from different models. Does the combination of several models of different complexity improve our capacity to synthesize different perceptions of the ecosystem functioning and therefore the value of the modelling enterprise in the context of ecosystem management? Our study addresses this question using a complex (14 state-variable) eutrophication model along with a simpler modelling construct that considers the interplay among phosphate, detritus, and generic phytoplankton and zooplankton state variables. Using Markov Chain Monte Carlo simulations, we calculate the relative mean standard error to assess the posterior support of the two models after considering the available data from the system. Predictions from the two models are then combined using the respective standard error estimates as weights in a weighted model average. The model averaging approach is used to examine the robustness of predictive statements made from our earlier work regarding the response of Hamilton Harbour (Ontario, Canada) to the different nutrient loading reduction strategies. In particular, we consolidate the finding that the existing total phosphorus goal (<17 $\mu g\,L^{-1}$) is most likely unattainable, and therefore we identify the most achievable ambient target under the most stringent (but realistic) nutrient loading reduction scenario. Finally, the discrepancy between the chlorophyll *a* predictions of the two models pinpoint the need to delve into the dynamics of phosphorus in the sediment–water column interface, as the internal nutrient loading can conceivably be a regulatory factor of the duration of the transient phase and the recovery resilience of the system.

> "...Because the goal of predictive limnology is predictive power, its advance is measured by decreased uncertainty of prediction and its controversies are resolved by comparing the ability of alternative approaches and theories to make the required prediction..." R.H. Peters (1986)

1. Introduction

Despite the considerable progress accomplished over the last 3–4 decades, the credibility of aquatic biogeochemical models to form the basis of public policy decisions has been severely criticized in the literature (Arhonditsis and Brett, 2004; Anderson, 2005; Flynn, 2005; Arhonditsis et al., 2006). With over 800 citations, the Oreskes et al. (1994) paper stands out as one of the classic critiques of the veracity of scientific methodology of earth sciences models, advocating the provocative standpoint that the validation of any type of model aiming to reproduce an open system is practically impossible. If we go beyond the controversy arising from the technical/philosophical meaning of validation (Rykiel, 1996), this statement essentially suggests that the inherent complexity of open environmental systems is subject to multiple conceptualizations and ever-changing mathematical descriptions. Being primarily a reflection of our current level of understanding and existing measurement technologies, the fact that many different model structures and many different parameter sets within a chosen model structure can acceptably reproduce the observed behaviour of a complex environmental system has long been discussed in the literature (Gauch, 1993; Beven and Freer, 2001; Christakos, 2002, 2003). Yet, Neuman (2003) pinpointed the

* Corresponding author. Tel.: +1 416 208 4858; fax: +1 416 287 7279.
E-mail address: georgea@utsc.utoronto.ca (G.B. Arhonditsis).

0304-3800/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ecolmodel.2012.05.023

absence of rigorous methodological frameworks to develop alternative site-specific conceptual mathematical models, select the optimal subset of models, effectively combine them to optimize predictions, and subsequently assess the underlying uncertainty.

Surprisingly, this very important notion is still somewhat neglected in the modelling literature, although there are viewpoints suggesting that environmental management decisions relying upon a single inadequate model can introduce bias and uncertainty that is much larger than the error stemming from inadequate choice of model parameter values (Neuman, 2003). In particular, the practise of basing ecological forecasts on one single mathematical model implies that a valid alternative model may be rejected (or omitted) from the decision making process (Type I model error), but also that our projections can be potentially the result of an erroneous mathematical construct that we failed to reject in an earlier stage (Type II model error). Further, the outputs of any model should be viewed through the prism of the underlying assumptions and therefore model acceptance in one or more settings is not evidence for general model applicability, but rather the start of a perpetual race for confirmation. The greater the number of cases in space or time in which the model is tested and confirmed, the higher the likelihood that its structure and conceptualization are still adequate to capture a substantial portion of the observed variability (Arhonditsis, 2009).

Recognizing that there is no true model of an ecological system, but rather several adequate descriptions of different conceptual basis and structure (Reichert and Omlin, 1997), Bayesian Model Averaging (BMA) is a technique designed to explicitly account for the uncertainty inherent in the model selection process. By averaging over many different competing models, BMA incorporates the uncertainty about the optimal model for any given exercise into the inference drawn about parameters and prediction (Raftery et al., 2005). Therefore, rather than picking the single "best-fit" model to predict future system responses, we can use Bayesian model averaging to provide a weighted average of the forecasts from different models (Hoeting et al., 1999). BMA has been applied successfully to many model classes including linear regression, generalized linear, exponential decay, discrete graphical, and dynamic models (Lamon and Clyde, 2000; Stow et al., 2004; Sloughter et al., 2007; Azim et al., 2011). In weather forecasting, BMA has offered a means for statistical post-processing of ensemble outputs, thereby achieving lower predictive error and sharper predictive probability density functions (Raftery et al., 2003; Bao et al., 2010; Sloughter et al., 2010). Further, the explicit consideration of both between- and within-forecast variability offered a meaningful explanation of the tendency of ensembles to exhibit significant spread-skill relationships, in which the spread in the ensemble forecasts is correlated with the magnitude of the forecast error, but yet to be underdispersive and thus uncalibrated (Raftery et al., 2003).

In the context of eutrophication, Lamon and Clyde (2000) used BMA to account for the uncertainty associated with the selection of the subset of explanatory variables (total phosphorous, total nitrogen, lake level, water temperature, wind speed and direction) and their optimal functional forms (i.e., linear predictors, regression spline predictors, or product spline interactions) to predict chlorophyll *a* concentrations in Lake Okeechobee. In a similar manner, BMA significantly improved the mean squared error for overall lake predictions, while the corresponding uncertainty bounds provided better coverage for new observations relative to the confidence intervals obtained by ordinary least squares models. Since the Lamon and Clyde (2000) pioneering work, notwithstanding the handful of statistical and process-based eutrophication modelling studies that have adopted Bayesian inference techniques (Borsuk et al., 2004; Arhonditsis et al., 2003, 2007; Qian and Reckhow, 2007; Scavia and Liu, 2009), there is an overwhelming gap in the literature of BMA approaches to guide eutrophication risk assessment. Moreover, there has been little focus on the benefits of basing ecological forecasts on combinations of process-based models, and practically no discussion on the ways through which the outputs of mathematical models with multiple endpoints (state variables, process rates) can be objectively integrated into one single averaged prediction.

To this end, the present study examines the potential benefits for model-based environmental management when a combination of models of different complexity is being used. Our study addresses this question using a complex eutrophication model, developed by Ramin et al. (2011) to guide the water quality criteria setting process in Hamilton Harbour (Ontario, Canada), along with a simpler plankton model that is founded upon the interplay among phosphate, detritus, and generic phytoplankton and zooplankton state variables. In this study, we first use a Bayesian framework to independently update the two ecological models with a dataset that represents the water quality conditions currently prevailing in Hamilton Harbour. We then weight the performance of the individual models into an average water quality prediction, which is used to examine the robustness of predictive statements made from our earlier work regarding the response of Hamilton Harbour to the different nutrient loading reduction strategies. Given the idiosyncrasies of ecological process-based modelling, our contention is that the Bayesian model average approach should not be viewed solely as a means to improve predictive capacity, but rather as an opportunity to compare alternative ecological structures and to integrate across different paradigms. We therefore compare the ecosystem dynamics postulated by the parameterization of the two models and identify ecological mechanisms that could potentially modulate the efficiency of the restoration efforts in the Hamilton Harbour area. Finally, we emphasize that future research should focus on the refinement of the weighting schemes and other performance standards to impartially synthesize the predictions of different models.

2. Methods

2.1. Study site

Hamilton Harbour is a large embayment at the western end of Lake Ontario and has a history of eutrophication problems manifested as algal blooms, low water transparency, prevalence of toxic cyanobacteria, and low hypolimnetic oxygen concentrations during the late summer (Hiriart-Baer et al., 2009; Ramin et al., 2011). Hamilton Harbour's drainage basin is about 500 km^2 in aerial extent and consists of watersheds dominated by agricultural land use (Grindstone and Spencer Creeks) and urban land use (Redhill and Indian Creeks; Fig. 1). The Harbour was the focus of intensive monitoring and modelling studies throughout the 70s and early 80s (e.g., Kohli, 1979; Klapwijk and Snodgrass, 1985). Since the mid 80s, when the Harbour was identified as one of the 43 Areas of Concern (*AOC*) by the Water Quality Board of the International Joint Commission, the remedial measures proposed by the Hamilton Harbour Remedial Action Plan (*RAP*) were based on the premise that the algal biomass levels and water clarity could be controlled by reducing ambient phosphorus concentrations. The substantial reduction of total phosphorus (*TP*) from the sewage effluents of the four wastewater treatment plants (*WWTPs*) and the steel mills that discharge into Hamilton Harbour, led to a significant decrease of *TP* concentrations and to an improvement of the water clarity, which in turn has triggered aquatic macrophyte resurgence in some nearshore areas. Yet, the system still receives substantial loads of phosphorus, ammonia, and suspended solids from the Burlington and Hamilton *WWTPs*, and arguably remains quite far from attaining the delisting *RAP* water quality goals (*TP* < 17 $\mu g\,L^{-1}$, chl *a* 5–10 $\mu g\,L^{-1}$, Secchi

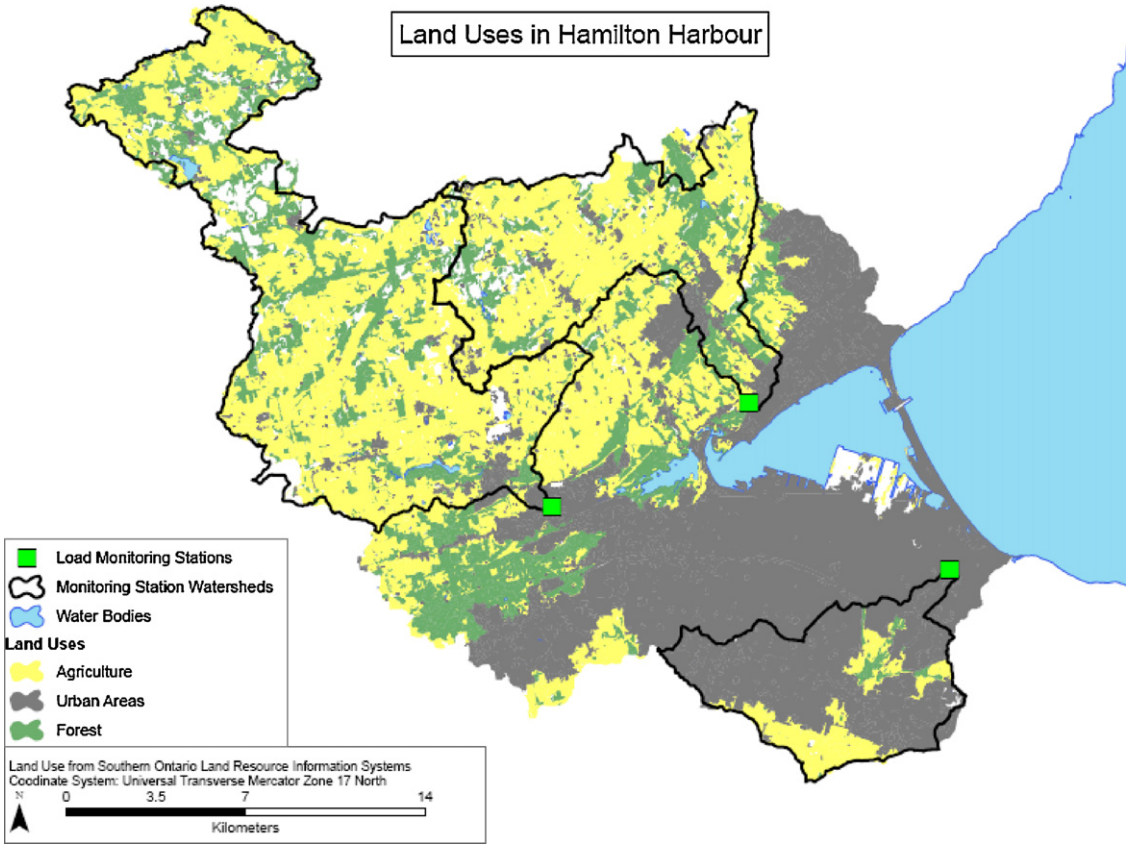


Fig. 1. Map of land uses in Hamilton Harbour watershed, western end of Lake Ontario, Ontario, Canada.

Disk Transparency > 3 m) set by the Stage 2 Update 2002 Report (Hamilton Harbour RAP Stakeholders, 2003).

Environmental modelling has been an indispensable tool of the Hamilton Harbour restoration efforts, and recent modelling work suggests the attainment of the water quality goal related to the summer chlorophyll *a* concentrations (5–10 $\mu g\,L^{-1}$) is achievable (Gudimov et al., 2010; Ramin et al., 2011). Yet, it has also been emphasized that the criteria setting process should explicitly accommodate the natural variability of the system by permitting a realistic frequency of goal violations. Namely, an exceedance frequency of 10% or less of the samples collected on a weekly basis during the focal period (i.e., June–September) should still be considered as compliance with the water quality goals of the system. A follow-up study by Gudimov et al. (2011) showed that the current epilimnetic total phosphorus target of 17 $\mu g\,L^{-1}$ is probably stringent and therefore a somewhat higher value (e.g., 20 $\mu g\,L^{-1}$) may provide a more realistic target. The same modelling analysis focused on the major ecological mechanisms that can potentially modulate the response of the system, and therefore shape the restoration rate as well as the stability of the new trophic state of the Harbour. It was suggested that the dynamics of phosphorus in the sediment–water column interface need to be revisited, as the internal nutrient loading can conceivably be a regulatory factor of the duration of the transient phase and the recovery resilience of the Harbour. Gudimov et al. (2011) also pinpointed two critical aspects of the system dynamics that invite further investigation and will likely determine our predictive capacity to assess compliance with the chlorophyll *a* criterion of 10 $\mu g\,L^{-1}$: (i) the importance of the nutrient recycling mediated by the microbial food web; and (ii) the ecological mechanisms that favour structural shifts towards a zooplankton community dominated by large-sized and fast-growing herbivores.

2.2. Model description

The basic conceptual design along with the key features of the two models used are provided in the Supporting Information section, while their detailed description can be found elsewhere (Arhonditsis and Brett, 2005; Arhonditsis et al., 2007; Law et al., 2009; Ramin et al., 2011). The complex eutrophication model considers the interactions among the following eight state variables in the water column: nitrate (NO_3), ammonium (NH_4), phosphate (PO_4), generic phytoplankton (*PHYT*), cyanobacteria-like phytoplankton (*CYA*), zooplankton, organic nitrogen (*ON*) and organic phosphorus (*OP*) (Figs. 2a and b). With this model, we considered a two-compartment vertical segmentation representing the epilimnion and hypolimnion of the Harbour. The depths of the two boxes varied with time and were explicitly defined based on extensive field measurements for the study period 1987–2007 (Dermott et al., 2007; Hiriart-Baer et al., 2009). We also developed a simple model that considers the interplay among the limiting nutrient (phosphate), phytoplankton, zooplankton, and detritus (particulate phosphorus); also known as NPZD model in the literature (Fig. 2c). The spatial segmentation of the model consists of three compartments, representing the epilimnion, mesolimnion, and hypolimnion of the system.

Thus, the present study selected two models at both ends of the complexity spectrum that have different strengths and weaknesses. One model is a simple mathematical description of the system that accounts for the interplay between the limiting nutrient and aggregated biotic compartments such as “phytoplankton”, and “zooplankton” (Edwards, 2001; Arhonditsis et al., 2007). This simple approach is more easily subjected to detailed uncertainty analysis and also has the advantage of fewer unconstrained parameters. The second model simulates two elemental cycles, functional

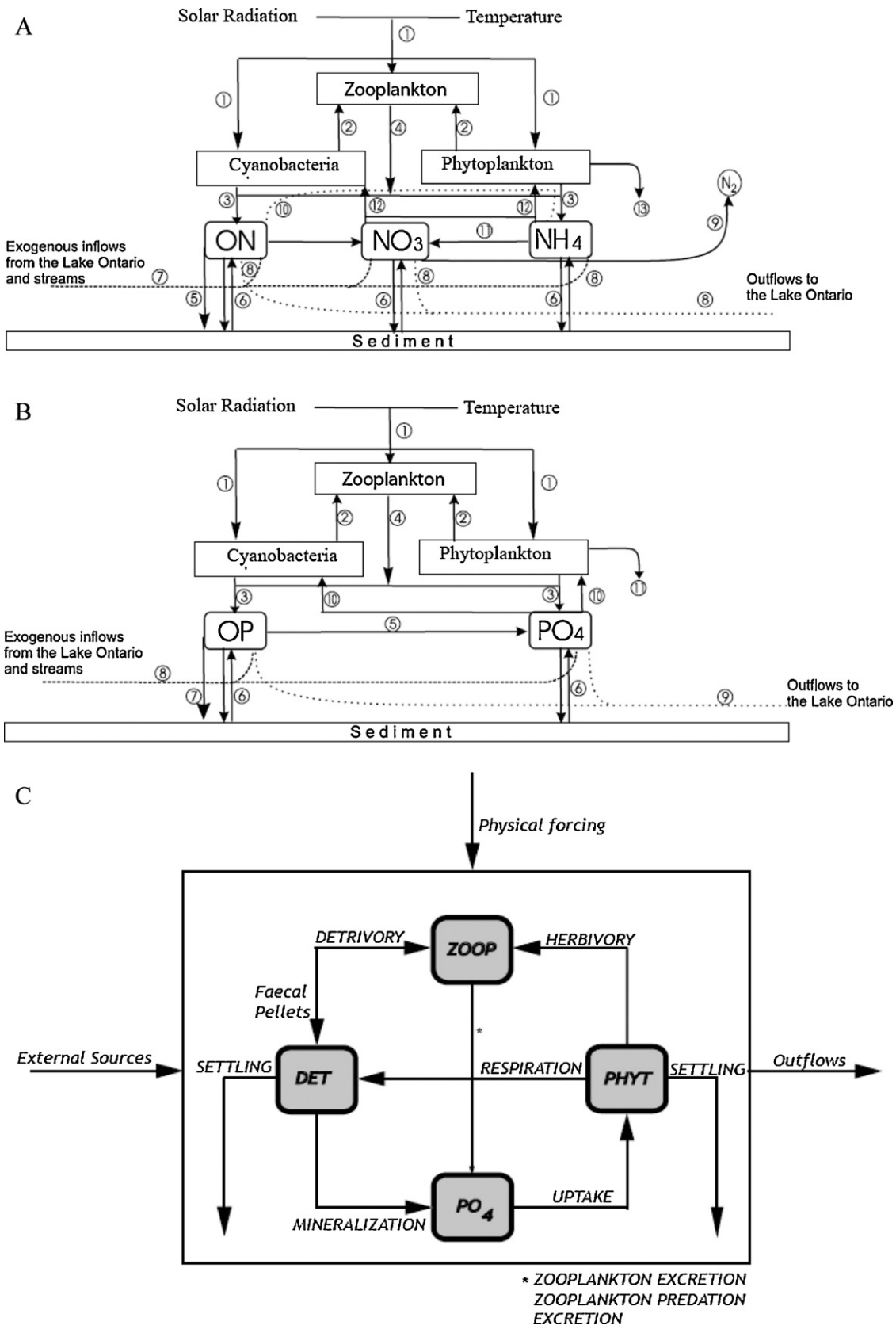


Fig. 2. (A) The nitrogen biogeochemical cycle of the model: (1) external forcing to phytoplankton growth (temperature, solar radiation); (2) zooplankton grazing; (3) phytoplankton basal metabolism excreted as NH_4 and *ON*; (4) zooplankton basal metabolism excreted as NH_4 and *ON*; (5) settling of particles; (6) water sediment NO_3, NH_4, and *ON* exchanges; (7) exogenous inflows of NO_3, NH_4, and *ON*; (8) outflows of NO_3, NH_4, and *ON*; (9) NO_3 sinks due to denitrification; (10) *ON* mineralization; (11) nitrification; and (12) phytoplankton uptake. *(B)* The phosphorus biogeochemical cycle of the model: (1) external forcing to phytoplankton growth (temperature, solar radiation); (2) zooplankton grazing; (3) phytoplankton basal metabolism excreted as PO_4 and *OP*; (4) zooplankton basal metabolism excreted as PO_4 and *OP*; (5) *OP* mineralization; (6) water sediment PO_4 and *OP* exchanges; (7) settling of particles; (8) exogenous inflows of PO_4 and *OP*; and (9) outflows of PO_4 and *OP*. *(C)* The flow diagram of the phosphate (PO_4) – phytoplankton (*PHYT*) – zooplankton (*ZOOP*) – detritus or particulate phosphorus (*DET*), also referred to as *NPZD* model.

phytoplankton groups, and dynamic nutrient release from the sediments. The sophisticated parameterization of the complex model provides confidence for more realistic reproduction of natural system dynamics, but the main criticism for this strategy is the inevitably poor identifiability with respect to the available data as well as the limited flexibility (high computational demands) to thoroughly examine model uncertainty to the input requirements. While all the predictive statements about the response of the system will be based on the synthesis of two models, our intent is also to highlight missing ecological processes, verify/reduce present model structures, and subsequently guide future model reformulation.

2.3. Bayesian framework

2.3.1. Statistical formulation:

We used a statistical formulation founded upon the assumption that the two eutrophication models are imperfect simulators of the environmental system and the corresponding process errors are invariant with the input conditions, i.e., the difference between each model and system dynamics was assumed to be constant over the annual cycle for each state variable. We also accounted for the uncertainty associated with the dataset using a data quality submodel. This component of our framework postulates that each observation from the system is a random draw from a normal distribution, in which the mean value represents the (latent) error-free observation (also referred to as "true value") and the variance is associated with the sampling error or other sources of uncertainty (e.g., variability in time/space). An observation i for the state variable j, yij, can be described as:

$$\begin{aligned} &y_{ij} \sim N(\hat{y}_{ij}, \sigma^2_{ijobs}) \\ &\hat{y}_{ij} \sim N(f(\theta, x_i, y_0), \sigma^2_j) \\ &\theta \sim MLN(\theta_\mu, \Sigma_\theta) y_0 \sim MN(y_{0\mu}, \Sigma_0) \\ &\sigma^2_{ijobs} = (0.25 \cdot y_{ij})^2 1/\sigma^2_j \sim gamma(0.001, 0.001) \\ &i = 1, 2, 3, \ldots, n \quad \text{and} \quad j = 1, \ldots, m \end{aligned} \tag{1}$$

where $\hat{y}_{ij}$ represents the latent "true value" used to parameterize the process-based models; σ_{ijobs} corresponds to the observation error; $f(\theta, x_i, y_0)$ denotes any of the two eutrophication models; σ_j is the time-independent, variable-specific process (structural) error term; x_i is a vector of time-dependent control variables (e.g., boundary conditions, forcing functions) describing the contemporary environmental conditions; the vector θ is a time independent set of the calibration model parameters; θ_μ indicates the vector of the mean values of θ in logarithmic scale; $\Sigma_\theta = I_\kappa \sigma_\theta{}^T \sigma_\theta$ and $\sigma_\theta = [\sigma_{\theta 1}, \ldots, \sigma_{\theta\kappa}]^T$ corresponds to the vector of the shape parameters of the κ lognormal distributions (standard deviation of $\log\theta$), where $\kappa = 16$ and 36 for the simple and complex models, respectively; y_0 corresponds to the vector of the values of the state variables at the initial time point t_0 (initial conditions); the vector $y_{0\,\mu} = [y_{1,1}, \ldots, y_{1,\pi}]^T$ corresponds to the January values of all the state variables of the model; *MLN* and *MN* represent the multivariate lognormal and multivariate normal distributions, respectively; $\pi = 12$ (simple model: 4 state variables $\times$ 3 spatial compartments) or 28 (complex model: 14 state variables $\times$ 2 spatial compartments). Depending on the available information from the system, the characterization of the prior density of their initial values was based on the assumption of a Gaussian distribution with mean values derived from the January monthly averages and moderately diffuse standard deviations, specified as 25% of the mean value for each state variable j; i.e., $\Sigma_0 = I_\pi (0.25)^2 y_{0\mu}{}^T y_{0\mu}$. Notably, when flat priors were assigned to the initial conditions, the inference drawn remained practically unaltered, although the posterior uncertainty associated with the initial conditions was somewhat greater. Similar to the practice followed by Ramin et al. (2011), our model calibration aimed to reproduce the average seasonal plankton dynamics in the current state of the system. [Description of the calibration dataset is also provided in Section 2 of the Supporting Information.]

2.3.2. Prior parameter distributions–numerical approximations for posterior distributions

The calibration vectors of the simple and complex eutrophication models consist of sixteen (16) and thirty six (36) parameters, respectively. Both vectors comprised the most influential parameters, as identified from earlier sensitivity analyses of the two models (Arhonditsis et al., 2007; Ramin et al., 2011). The prior parameter distributions reflected the existing knowledge (field observations, laboratory studies, literature information and expert judgment) on the relative plausibility of their values. The characterization of the parameter distributions was similar to the protocol introduced by Steinberg et al. (1997) and subsequently used in several recent studies (Arhonditsis et al., 2007, 2008a,b; Zhang and Arhonditsis, 2008; Ramin et al., 2011). Namely, we identified the minimum and maximum values for each parameter and then we assigned lognormal distributions parameterized such that 95% of their values (the two tail areas had equal probability mass) lay within the identified ranges. The prior distributions of all the calibration parameters of the two models are presented in Tables 1 and 2. [The default values assigned to the parameters of the complex model that were not considered during the updating exercise are provided in the Supporting Information section.]

Sequence of realizations from the posterior distribution of the model was obtained using Markov chain Monte Carlo (MCMC) simulations (Gilks et al., 1998). We used the general normal-proposal Metropolis algorithm coupled with an ordered over-relaxation to control the serial correlation of the MCMC samples (Neal, 1998). As originally proposed by Arhonditsis et al. (2007), the present Bayesian parameter estimation is based on two parallel chains with starting points: (i) a vector that consists of the mean values of the prior parameter distributions, and (ii) a vector based on a preliminary deterministic calibration of the two models. The models were run for 30,000 iterations and convergence was assessed with the modified Gelman–Rubin convergence statistic (Brooks and Gelman, 1998). The accuracy of the posterior parameter values was inspected by assuring that the Monte Carlo error for all the parameters was less than 5% of the sample standard deviation. Our framework is implemented in the WinBUGS Differential Interface (WBDiff); an interface that allows numerical solution of systems of ordinary differential equations within the WinBUGS software.

Aside from the differences in the central tendency and the underlying uncertainty, we also evaluated the degree of updating between parameter priors and posteriors by assessing the changes in the shape of the corresponding distributions using the delta index (Endres and Schindelin, 2003; Hong et al., 2005). The delta index measures the distance between two probability distributions:

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

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

Table 1
Parameter definitions and Markov Chain Monte Carlo posterior estimates of the mean values and standard deviations of the stochastic nodes of the simple eutrophication model.

Parameters	Description	Units	Priors		Posteriors		References
			Mean	SD	Mean	SD	
a	Maximum phytoplankton growth rate	day^{-1}	1.772	0.382	2.134	0.251	Gudimov et al. (2010)
d	Zooplankton mortality rate	day^{-1}	0.114	0.015	0.110	0.009	Lampert and Sommer (1997), Omlin et al. (2001), Sommer (1989), Jorgensen et al. (1991), Chen et al. (2002, and references therein)
K_p	Half-saturation constant for PO_4 uptake	mg P m^{-3}	13.011	4.666	10.63	2.057	Cerco and Cole (1993, and references therein), Arhonditsis and Brett (2005), Reynolds (2006)
r	Phytoplankton respiration rate	day^{-1}	0.035	0.016	0.090	0.011	Edwards (2001)
s	Phytoplankton sinking loss rate	m day^{-1}	0.068	0.050	0.046	0.018	Edwards (2001)
μ	Zooplankton grazing half-saturation coefficient	mg P m^{-3}	5.296	3.359	16.09	2.189	Sommer (1989), Jorgensen et al. (1991)
φ	Detritus mineralization rate	day^{-1}	0.044	0.025	0.013	0.002	Gudimov et al. (2010)
ψ	Detritus sinking rate	m day^{-1}	0.341	0.252	0.443	0.012	Edwards (2001)
λ	Maximum zooplankton grazing rate	day^{-1}	0.571	0.077	0.582	0.049	Sommer (1989), Jorgensen et al. (1991)
K_b	Background light extinction coefficient	m^{-1}	0.214	0.029	0.204	0.018	Hamilton and Schladow (1997)
K_c	Light extinction coefficient due to chlorophyll a	L (μg *chla* m)$^{-1}$	0.031	0.013	0.023	0.005	Arhonditsis and Brett (2005), Hamilton and Schladow (1997)
a	Zooplankton assimilation efficiency	–	0.491	0.039	0.471	0.022	Gudimov et al. (2010)
β	Zooplankton excretion fraction to phosphate	–	0.385	0.095	0.292	0.048	Edwards (2001)
γ	Zooplankton predation excretion fraction to phosphate	–	0.385	0.095	0.445	0.071	Edwards (2001)
ω	Relative zooplankton preference for detritus compared to phytoplankton	–	0.470	0.150	0.237	0.062	Gudimov et al. (2010)
I_s	Half saturation light intensity	MJ m^{-2} day^{-1}	160.6	28.79	146.8	16.88	Gudimov et al. (2010)

Table 2
Parameter definitions and Markov Chain Monte Carlo posterior estimates of the mean values and standard deviations of the stochastic nodes of the complex eutrophication model.

Parameters	Description	Units	Priors Mean	SD	Posteriors Mean	SD	References
$AH_{(cy)}$	Half saturation constant for ammonium cyanobacteria uptake	$\mu g\ N\ L^{-1}$	49.88	9.585	50.48	9.575	Cerco and Cole (1993, and references therein)
$AH_{(phyt)}$	Half saturation constant for ammonium phytoplankton uptake	$\mu g\ N\ L^{-1}$	110.4	13.517	112.7	11.73	Cerco and Cole (1993, and references therein)
$a_{PO4(cy)}$	Fraction of cyanobacteria basal metabolism released as phosphate		0.385	0.095	0.390	0.098	Cerco and Cole (1993, and references therein)
$a_{PO4(phyt)}$	Fraction of phytoplankton basal metabolism released as phosphate		0.385	0.095	0.338	0.074	Cerco and Cole (1993, and references therein)
$a_{PO4(zoop)}$	Fraction of zooplankton basal metabolism released as phosphate		0.385	0.095	0.625	0.089	Cerco and Cole (1993, and references therein)
Denitrifmax rate	Maximum denitrification	$\mu g\ N\ L^{-1}\ day^{-1}$	3.494	1.643	4.038	2.036	Gudimov et al. (2010)
$Filter_{(cy)}$	Cyanobacteria filtering rate from dreissenids	day^{-1}	0.012	0.004	0.011	0.003	Gudimov et al. (2010)
$Filter_{(phyt)}$	Phytoplankton filtering rate from dreissenids	day^{-1}	0.023	0.008	0.015	0.003	Gudimov et al. (2010)
$Growthmax_{(cy)}$	Cyanobacteria maximum growth rate	day^{-1}	1.350	0.155	1.319	0.138	Gudimov et al. (2010)
$Growthmax_{(phyt)}$	Phytoplancton maximum growth rate	day^{-1}	2.574	0.155	2.649	0.153	Gudimov et al. (2010)
$Ik_{(cy)}$	Half saturation light intensity for cyanobacteria	$MJ\ m^{-2}\ day^{-1}$	160.6	28.79	162.0	27.97	Gudimov et al. (2010)
$Ik_{(phyt)}$	Half saturation light intensity for phytoplankton	$MJ\ m^{-2}\ day^{-1}$	160.6	28.79	120.9	13.89	Gudimov et al. (2010)
$k_{background}$	Background light extinction coefficient	m^{-1}	0.214	0.029	0.234	0.025	Hamilton and Schladow (1997, and references therein)
$kchla_{(cy)}$	Self shading effect for cyanobacteria	$L\ (\mu g\ chla\ m)^{-1}$	0.031	0.013	0.032	0.011	Arhonditsis and Brett (2005), Hamilton and Schladow (1997, and references therein)
$kchla_{(phyt)}$	Self shading effect for phytoplankton	$L\ (\mu g\ chla\ m)^{-1}$	0.026	0.009	0.070	0.007	Arhonditsis and Brett (2005), Hamilton and Schladow (1997, and references therein)
$KC_{refmineral}$	Organic carbon mineralization rate	day^{-1}	0.044	0.025	0.021	0.009	Gudimov et al. (2010)
$KN_{refmineral}$ rate	Nitrogen mineralization	day^{-1}	0.044	0.025	0.028	0.010	Cerco and Cole (1993, and references therein), Hamilton and Schladow (1997, and references therein)
$K_{Prefmineral}$ rate	Phosphorus mineralization	day^{-1}	0.044	0.025	0.011	0.002	Omlin et al. (2001), Hamilton and Schladow (1997, and references therein), Cerco and Cole (1993, and references therein)
K_Z	Half saturation constant for zooplankton grazing	$\mu g\ C\ L^{-1}$	98.28	7.747	96.16	6.556	Sommer (1989), Jorgensen et al. (1991)
max grazing	Zooplankton maximum grazing rate	day^{-1}	0.491	0.039	0.471	0.014	Sommer (1989), Jorgensen et al. (1991)

Table 2 (*Continued*)

Parameters	Description	Units	Priors Mean	SD	Posteriors Mean	SD	References
$mp_{(cy)}$ rate	Cyanobacteria mortality	day^{-1}	0.023	0.008	0.023	0.006	Omlin et al. (2001), Jorgensen et al. (1991), Cerco and Cole (1993, and references therein), Reynolds (1984)
$mp_{(phyt)}$ rate	Phytoplankton mortality	day^{-1}	0.023	0.008	0.027	0.006	Omlin et al. (2001), Jorgensen et al. (1991), Cerco and Cole (1993, and references therein), Reynolds (1984), Hamilton and Schladow (1997, and references therein)
mz rate	Zooplankton mortality	day^{-1}	0.156	0.015	0.148	0.005	Lampert and Sommer (1997), Omlin et al. (2001), Sommer (1989), Jorgensen et al. (1991), Chen et al. (2002, and references therein)
$NH_{(cy)}$	Half saturation constant for nitrate cyanobacteria uptake	$\mu g\ N\ L^{-1}$	49.88	9.585	49.92	9.714	Arhonditsis and Brett (2005), Hamilton and Schladow (1997, and references therein)
$NH_{(phyt)}$	Half saturation constant for nitrate phytoplankton uptake	$\mu g\ N\ L^{-1}$	110.4	13.52	110.2	13.54	Arhonditsis and Brett (2005), Hamilton and Schladow (1997, and references therein)
Nitrifmax rate	Maximum nitrification	$\mu g\ N\ L^{-1}.day^{-1}$	17.72	3.822	9.338	0.980	Reynolds (1984), Hamilton and Schladow (1997, and references therein), Berounsky and Nixon (1990)
$PH_{(cy)}$	Half saturation constant for phosphorus cyanobacteria uptake	$\mu g\ P\ L^{-1}$	23.35	2.321	23.87	2.298	Cerco and Cole (1993, and references therein), Arhonditsis and Brett (2005), Reynolds (2006)
$PH_{(phyt)}$	Half saturation constant for phosphorus phytoplankton uptake	$\mu g\ P\ L^{-1}$	8.859	1.911	8.541	1.551	Cerco and Cole (1993, and references therein), Arhonditsis and Brett (2005), Reynolds (2006)
$P_{maxuptake(cy)}$	Maximum phosphorus uptake rate for cyanobacteria	$\mu g\ P\ L^{-1}.\ day^{-1}$	0.016	0.003	0.014	0.002	Jorgensen et al. (1991), Arhonditsis and Brett (2005), Hamilton and Schladow (1997, and references therein)
$P_{maxuptake(phyt)}$	Maximum phytoplankton uptake rate for phytoplankton	$\mu g\ P\ L^{-1}.\ day^{-1}$	0.023	0.008	0.019	0.003	Jorgensen et al. (1991), Arhonditsis and Brett (2005), Hamilton and Schladow (1997, and references therein)
$V_{settling}$	Allochthonous particle settling velocity	$m\ day^{-1}$	0.886	0.191	0.395	0.035	Chen et al., 2002 (and references therein), Cerco and Cole (1993, and references therein), Arhonditsis and Brett (2005), Reynolds (2006)
$V_{settling(biogenic)}$	Biogenic particle settling	$m\ day^{-1}$	0.117	0.038	0.102	0.029	Gudimov et al. (2010)
$V_{settling(cy)}$	Cyanobacteria settling	$m\ day^{-1}$	0.023	0.008	0.023	0.007	Wetzel (2001),Cerco and Cole (1993, and references therein), Sandgren (1991)
$V_{settling(phyt)}$	Phytoplankton settling	$m\ day^{-1}$	0.177	0.038	0.184	0.031	Wetzel (2001),Cerco and Cole (1993, and references therein), Sandgren (1991)
β_N	Fraction of inert nitrogen buried into deeper sediment		0.428	0.058	0.539	0.097	Gudimov et al. (2010)
β_P	Fraction of inert phosphorus buried into deeper sediment		0.845	0.039	0.918	0.031	Gudimov et al. (2010)

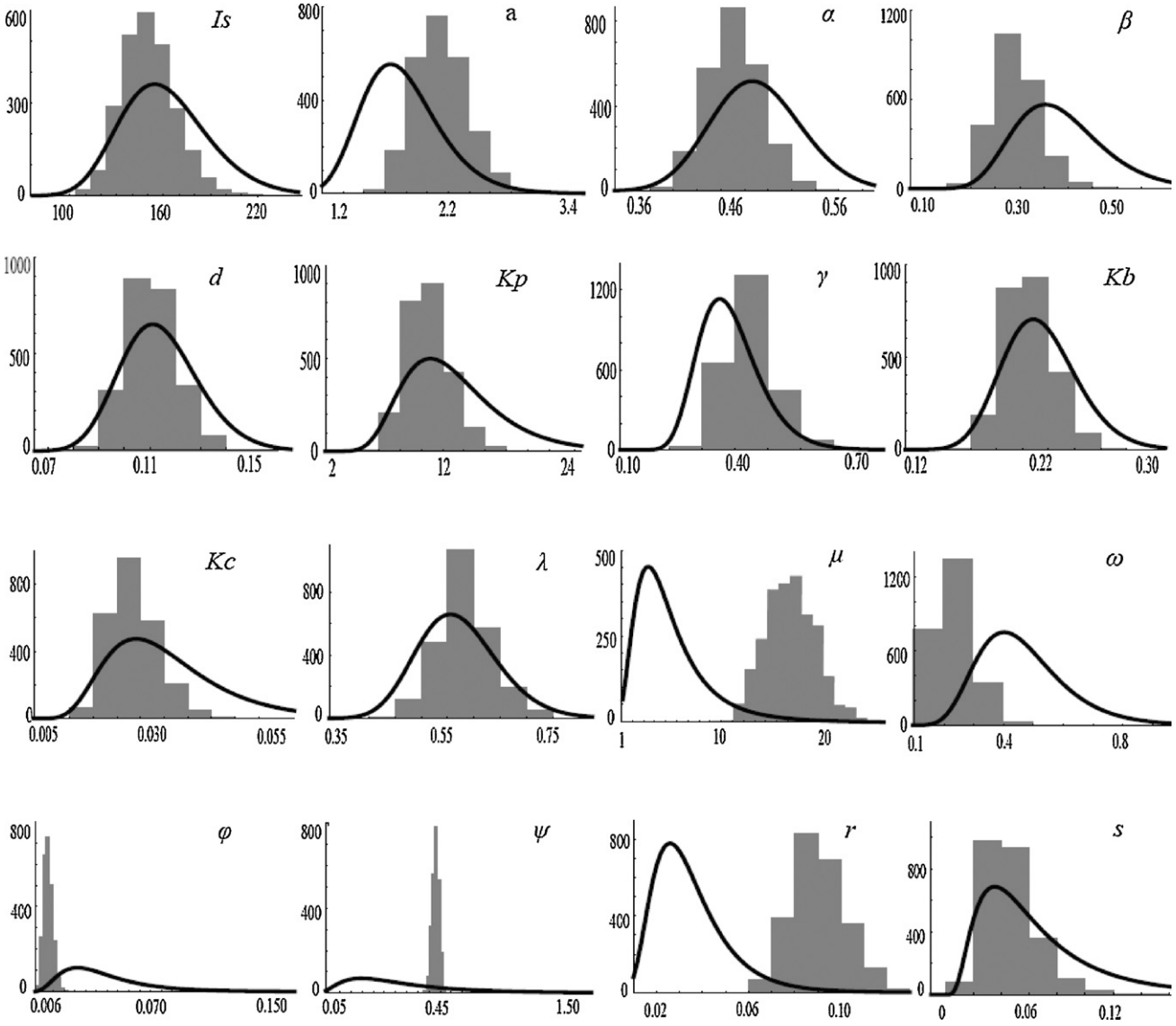


Fig. 3. Comparison between the prior (black line) and posterior parameter distributions of the calibration vector of the simple model for Hamilton Harbour. The *Y* axis represent the number of Markov chain Monte Carlo samples included within each bin of the corresponding parameter spaces.

2.3.3. Model updating and loading scenarios

The MCMC estimates of the mean and standard deviation parameter values along with the covariance structure were used to update the two models (Gelman et al., 1995). Under the assumption of a multivariate normal distribution for the parameter values, the conditional distributions are given by:

$$\hat{\theta}_{i|j} = \hat{\theta}_i + [\theta_j - \hat{\theta}_j]\Sigma_j^{-1}\Sigma_{i,j} \tag{3}$$

$$\Sigma_{i|j} = \Sigma_i - \Sigma_{j,i}\Sigma_j^{-1}\Sigma_{i,j} \quad j \in \{i+1, \ldots n\} \tag{4}$$

where $\hat{\theta}_{i|j}$ and $\Sigma_{i|j}$ correspond to the mean value and the dispersion matrix of the parameter i conditional on the parameter vector j; the values of the elements Σ_i, $\Sigma_{i,j}$, and Σ_j correspond to the variance and covariance of the two subset of parameters; and $\hat{\theta}_i$, $\hat{\theta}_j$, θ_j correspond to the posterior mean and random values of the parameters i and j, respectively. The updated models provided the basis for a series of posterior simulations that aimed to examine the compliance of the system with the targeted water quality standards, if the Hamilton Harbour RAP loading recommendations are actually implemented. Summary statistics of the exogenous flows and nutrient loadings used to force the models were similar to those presented by Ramin et al. (2011) (see Table 6 in their Electronic Supplementary Material). We also combined the predictions from the two models using the respective mean standard error estimates as weights in a weighted model average:

$$w_{ij} = \frac{\sum_{k=1}^{MC} \frac{\sigma_{ijk}}{\bar{Y}_j}}{MC} \tag{5}$$

$$w_{Mi} = \frac{m}{\sum_{j=1}^{m} w_{ij}} \tag{6}$$

$$\overline{TP} = \sum_{i=1}^{l} w_{Mi} TP_{Mi} \quad \overline{chla} = \sum_{i=1}^{l} w_{Mi} chla_{Mi} \tag{7}$$

where l represents the number of models considered in this analysis ($l=2$); m corresponds to the number of state variables j of the model M_i for which data are available ($m=6$ or 11); MC is the total number of MCMC runs sampled to form the model posteriors; σ_{ijk} denotes the model structural error for the state variable j of the model M_i as sampled from the MCMC run k; $\bar{Y}_j$ represents the annual observed average for the variable j, TP_{Mi} and $chla_{Mi}$ are the

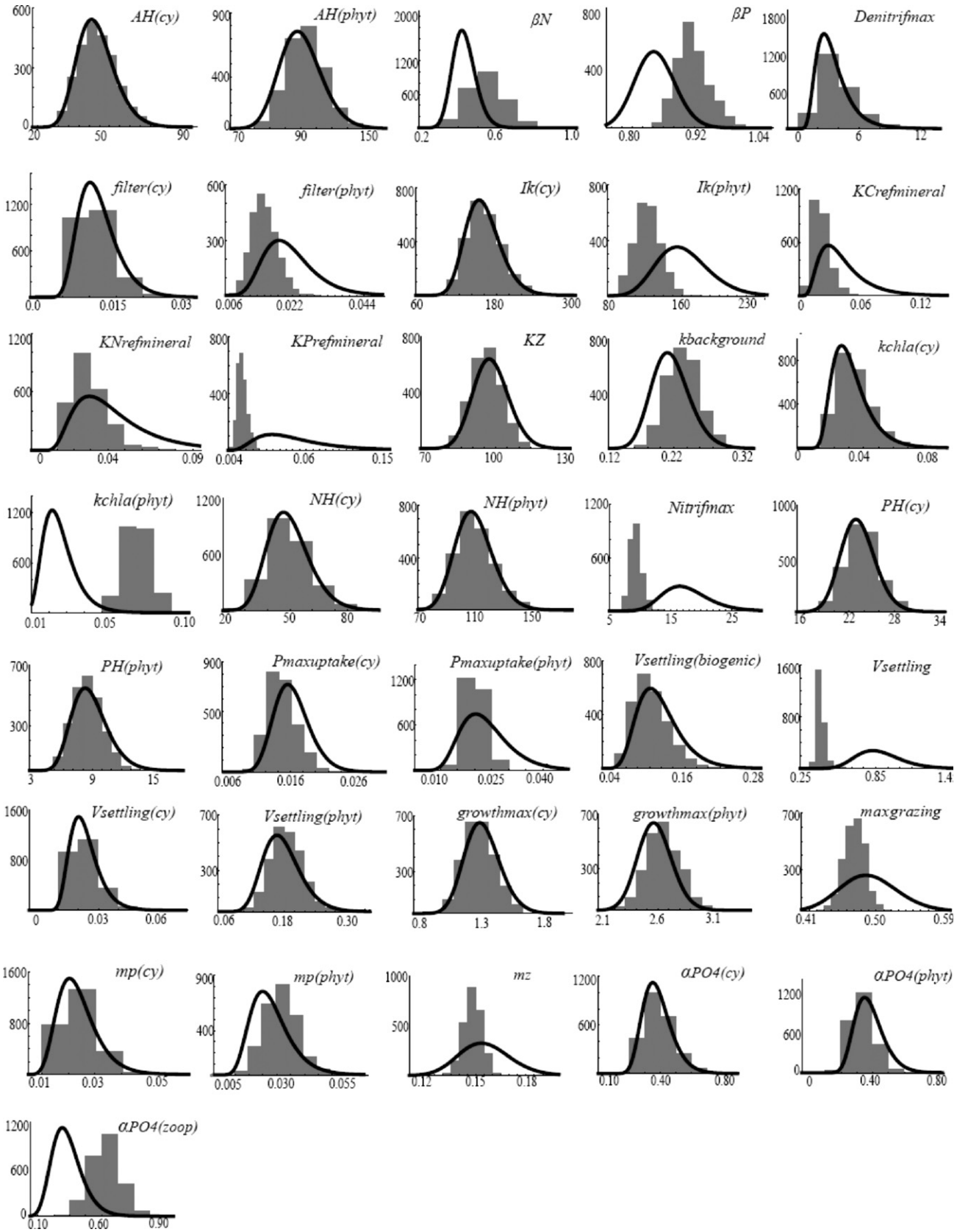


Fig. 4. Comparison between the prior (black line) and posterior parameter distributions of the calibration vector of the complex eutrophication model for Hamilton Harbour. The Y axes represent the number of Markov chain Monte Carlo samples included within each bin of the corresponding parameter spaces.

total phosphorus and chlorophyll a predictions from the individual models weighted by the corresponding weights w_{Mi} to obtain the averaged predictions $\overline{TP}$ and $\overline{chla}$.

3. Result

The two MCMC sequences for each of the two models converged rapidly (≈5000 iterations) and the statistics reported were based on the last 25,000 draws by keeping every 20th iteration (thin = 20). Mean values and standard deviations of the sixteen parameters of the simple model are provided in Table 1 and Fig. 3. Similarly, the central tendencies of the thirty six marginal parameter posterior distributions of the complex eutrophication model along with the underlying uncertainty are presented in Table 2 and Fig. 4. Relative to the prior parameter specification, the posterior statistics generally indicate that a substantial amount of knowledge was gained after updating the simple model. Characteristic examples were the substantial shifts of the most likely values of the zooplankton half saturation constant (μ), respiration rate (r), detritus mineralization rate (φ), and the zooplankton preference for detritus relative to phytoplankton (ω). We also note the significantly narrower posterior standard deviations of the mineralization rate (φ), detritus sinking rate (ψ), phytoplankton settling rate (s), phosphorus half saturation constant (K_p), and light extinction coefficient due to chlorophyll a (K_c) compared to those specified prior to the calibration. Likewise, several of the marginal posteriors of the complex model were characterized by significant shifts of their central tendency relative to the prior assigned values, e.g., self shading effect for phytoplankton ($kchla_{(phyt)}$), carbon/phosphorus mineralization rates ($KC_{refmineral}$, $KP_{refmineral}$), maximum nitrification rate ($Nitrifmax$), fraction of zooplankton basal metabolism released as phosphate ($a_{PO4(zoop)}$), and allochthonous particle settling velocity ($V_{settling}$). The same parameters along with the half saturation light intensity for phytoplankton ($Ik_{(phyt)}$), phytoplankton filtering rate from dreissenids ($filter_{(phyt)}$), zooplankton grazing ($max\ grazing$), and mortality rate (mz) demonstrated significant reductions of their posterior standard deviations. Interestingly, there were also several parameters of the complex model with posterior distributions that remained unaltered, suggesting that the dataset used did not offer any insights into the characterization of the cyanobacteria-like functional group ($Ik_{(cy)}$, $NH_{(cy)}$, $AH_{(cy)}$, $PH_{(cy)}$, $growthmax_{(cy)}$) as well as the phytoplankton nitrogen uptake kinetics ($AH_{(phyt)}$, $NH_{(phyt)}$).

Changes in the shape of the parameter distributions after the updating of the two eutrophication models are presented in Fig. 5. In the simple model, zooplankton grazing half-saturation coefficient (μ), mineralization rate (φ), detritus sinking rate (ψ) and phytoplankton respiration rate (r) showed the highest distance between priors and posteriors, whereas the zooplankton mortality rate (d), background light extinction coefficient (kb), and maximum zooplankton grazing rate (λ) demonstrated the lowest values. On the other hand, the self shading effect for phytoplankton ($kchla_{(phyt)}$), the fraction of inert nitrogen and phosphorus buried into deeper sediment (β_N, β_P), fraction of zooplankton basal metabolism released as phosphate ($a_{PO4(zoop)}$), the zooplankton mortality rate (mz), and the allochthonous particle settling velocity ($V_{settling}$) were characterized by a substantial shape change of their distributions after the calibration of the complex eutrophication model. Notably, the delta index values reiterate the previously reported finding that most of the parameters associated with the nitrogen utilization by the two phytoplankton functional groups ($AH_{(cy)}/_{(phyt)}$, $NH_{(cy)}/_{(phyt)}$), and the cyanobacteria specification remained practically the same, suggesting that whether the initial priors assigned were reasonable and/or that our knowledge about their values did not improve after the consideration of the dataset (Hong et al., 2005).

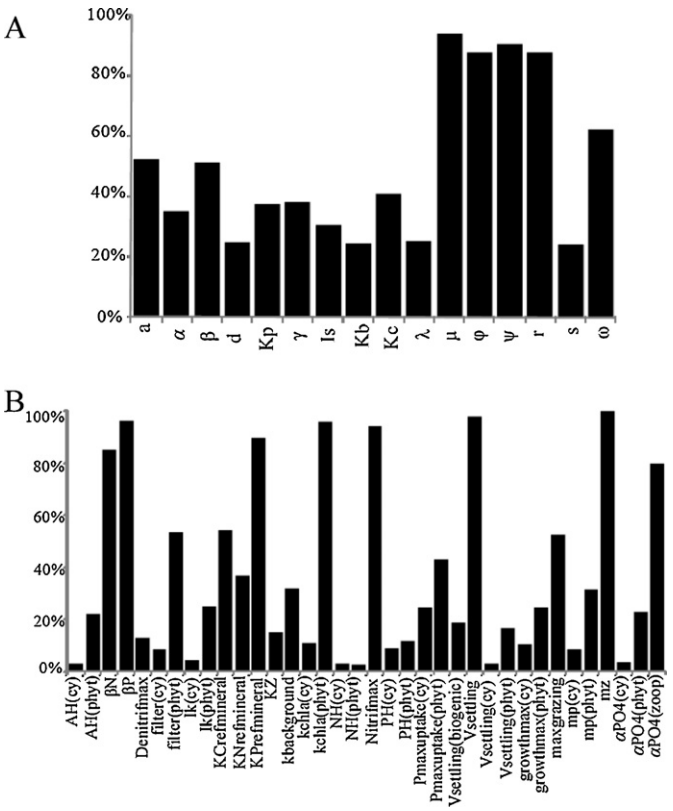


Fig. 5. Changes in the shape of the parameter distributions after the updating of the two eutrophication models. The assessment is based on the delta index (Eq. (2)) and all values are presented as percentages of the maximum value $\sqrt{2\log 2}$, in which there is no overlap between the two distributions.

The seasonally invariant model structure (or process) error terms (σ_j) delineate constants zones around the model endpoints for which empirical information from the system exists (Table 3). First, we highlight the notably lower error values of the complex model for nearly all the state variables, such as epilimnetic phosphate, phytoplankton and zooplankton biomass, hypolimnetic phosphate and organic phosphorus concentrations. Epilimnetic organic phosphorus is the only case for which the process error was somewhat higher with the complex model. Epilimnetic nitrate concentrations demonstrate the highest error value with the complex model (σ_{NO3epi} =240.5), whereas phytoplankton ($\sigma_{PHYTepi}$ = 205.5) is characterized by significant error with the simple model. By contrast, the simple and complex eutrophication models reproduce accurately the epilimnetic detritus ($\sigma_{DET/OPepi}$ = 1.083) and epilimnetic phosphate (σ_{PO4epi} = 0.287), respectively.

The posterior medians along with the 95% credible intervals derived from the calibration of the simple model were close to the observed values for epilimnetic phosphate, total phosphorus, chlorophyll a concentrations, total zooplankton biomass, and hypolimnetic phosphate in Hamilton Harbour (Fig. 1 in Supporting Information or Fig. SI-1). The model accurately predicts the epilimnetic phosphate levels including the winter maxima and summer minima. The simple model captures -although somewhat underestimated- the two major peaks of the zooplankton biomass, and closely represents the summer phytoplankton abundance (≈16 μg $chla$ L^{-1}). Yet, the chlorophyll a concentrations in the fall are distinctly over-predicted (>10 μg $chla$ L^{-1}), while the model fails to reproduce the phosphate hypolimnetic accumulation in the summer. Epilimnetic phosphate, total phosphorus, total ammonia-nitrogen, zooplankton biomass, hypolimnetic phosphate, and ammonia predictions from the complex model are

Table 3
Monte Carlo posterior estimates of the mean values and standard deviations of the model structural (or process) error terms of the two eutrophication models.

Parameters	NPZD			Complex		
	Mean	SD	RE[a]	Mean	SD	RE
σ_{PO4epi}	1.732	0.457	19.8%	0.287	0.101	2.6%
$\sigma_{PHYTepi}$	205.5	52.51	29.3%	–	–	–
$\sigma_{ZOOPepi}$	55.47	19.96	34.8%	13.86	3.466	7.7%
$\sigma_{DET/OPepi}$	1.083	0.421	2.7%	2.221	0.599	7.0%
σ_{NH4epi}	–	–	–	48.56	11.71	7.8%
σ_{NO3epi}	–	–	–	240.5	66.52	9.1%
σ_{CYAepi}	–	–	–	111.7	29.76	47.7%
$\sigma_{NONCYAepi}$	–	–	–	52.32	14.71	10.6%
$\sigma_{PO4hypo}$	2.261	0.533	22.2%	0.834	0.218	7.6%
$\sigma_{DET/OPhypo}$	3.794	0.898	13.9%	2.212	0.554	7.7%
$\sigma_{NH4hypo}$	–	–	–	14.81	5.33	2.2%
$\sigma_{NO3hypo}$	–	–	–	268.2	64.39	11.0%

[a] Relative Error (RE) = Σ|Observation − Prediction|/ΣObservation.

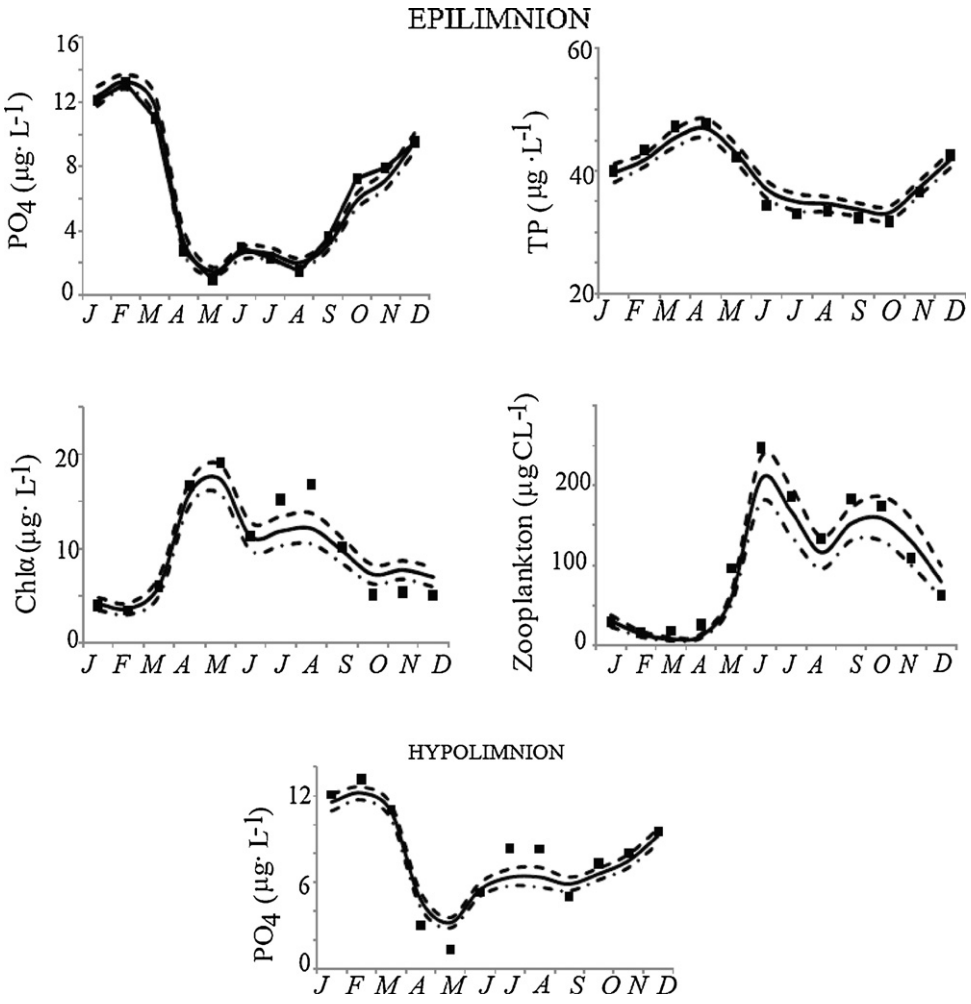


Fig. 6. Comparison between the observed data (black dots) and the averaged predictions of the two eutrophication models for total phosphorus, phosphate, chlorophyll *a*, and total zooplankton biomass in the Hamilton Harbour.

close to the observed data available from Hamilton Harbour (Fig. SI-2). The complex model accurately predicts the spring phytoplankton bloom (≈20 μg *chla* L^{-1}), but distinctly underpredicts the end of summer-mid fall secondary chlorophyll *a* peak (<13 μg *chla* L^{-1}); mainly due to underestimation of the abundance of the cyanobacteria-like functional group. Notably, nitrate concentrations appear to be overestimated in the epilimnion and underestimated in the hypolimnion. In the next step, the overall performance of the two models provided the basis for calculating weights (Eqs. (5)–(7)), which were then used to average their predictions. In particular, the posterior weights assigned to the complex and simple model were 0.63 and 0.37, respectively. The greater emphasis placed on the outputs from the complex model is clearly manifested by the improved representation of the hypolimnetic phosphate progressive increase during the stratified period, and the faithful description of the phytoplankton and zooplankton biomass seasonal cycles (Fig. 6). Yet, we note that the averaged prediction from the two models still underpredicts the summer

epilimnetic chlorophyll *a* levels, although the discrepancy is lower (≈1–2 μg *chla* L^{-1}) than the prediction supported by the complex eutrophication model alone.

Regarding the nutrient loading scenario examined, both the simple model (18.7 ± 0.7 μg *TP* L^{-1}) and the complex one (17.8 ± 0.9 μg *TP* L^{-1}) predict that the implementation of the HH RAP loading recommendations will reduce the average *TP* concentrations during the summer stratified period below the level of 20 μg *TP* L^{-1}, but the associated distributions will have a large portion of the probability mass above the Stage 2 water quality goal of a maximum of 17 μg *TP* L^{-1} (Fig. 7a and b). The complete agreement between the two forecasts for total phosphorus is also reflected in their averaged prediction (Fig. 7c). Both models also predict that the epilimnetic chlorophyll *a* concentrations will fall below the threshold level of 10 μg *chla* L^{-1} (Fig. 8a and b). Yet, the simple model appears to support more optimistic predictions with respect to phytoplankton response to the reduced ambient *TP* concentrations relative to the complex one. Consequently, the averaged predictive distribution for chlorophyll *a* demonstrates a distinct bimodal pattern with a primary mode at 7.5 μg *chla* L^{-1}, reflecting the greater weight placed on the complex model, and a secondary peak at 5.1 μg *chla* L^{-1}, associated with the simple one (Fig. 8c).

4. Discussion

An increasingly popular notion in the modelling literature is that the current generation of ecological models is far from having the capacity to address the complicated issues pertaining to the contemporary management of natural resources (Clark et al., 2001). Yet, the lack of a healthy dose of criticism and the tendency of modellers "to carry their trade far beyond the limits of reality" has undermined the trustworthiness of mathematical modelling to ultimately address vexing aspects of environmental stewardship (Anderson, 2005; Flynn, 2005; Pilkey and Pilkey-Jarvis, 2007). Indeed, despite the compelling reasons for identifying the idiosyncrasies and knowledge gaps of the natural environment, for differentiating between predictable and unpredictable patterns, and for critically evaluating model outputs, modellers tend to overstate the power of models and frequently make unfounded statements based on misleadingly deterministic projections. In many respects, the examination of uncertainty along with the objective recognition of what we can actually learn from the application of a mathematical model can be more insightful than its development, calibration or validation. In this regard, the present study focused on one profoundly underappreciated aspect of the model-based decision making process that involves the uncertainty arising from the adoption of one single model for a given environmental management problem. What is the additional knowledge gained by considering two or more mechanistic models and to what extent does the value of the modelling enterprise in the context of ecosystem management benefit from such a synthesis of predictions?

One of the critical decisions when considering models of different complexity involves the selection of the averaging scheme to synthesize their predictions. In this study, we opted for a strategy that considers performance over all the model endpoints rather than the subset of state variables included in both models or the variables more closely related to the environmental management problem at hand. While this approach may entail the risk of downplaying the impact of the best performing model for a particular variable, we believe that a fair assessment of the value of all the models integrated in an ensemble ecological forecast should strive for balanced performance over their entire structure. This approach aims to penalize the likelihood of calibration bias, whereby the maximization of the fit for a specific state variable (e.g., phytoplankton biomass, dissolved oxygen) may be accompanied by high error for other state variables (herbivorous zooplankton biomass, nutrient concentrations). In doing so, our intent is to avoid forecasts founded upon models with misleadingly high weights that conceal fundamentally flawed ecological structures (Franks, 1995; Arhonditsis and Brett, 2004). Future efforts to develop weighting schemes suitable for the synthesis of ecosystem model predictions will greatly benefit by several interesting statistical post-processing methods presented in the field of ensemble weather forecasting. For example, Wilks (2002) proposed the smoothing of forecast ensembles through fitting of mixtures of Gaussian distributions to ensemble data, and his results suggest that such smoothed ensembles improve probabilistic forecasting with small ensemble sizes and extreme events. Other simple statistical approaches with promising results involve the direct adjustment of the probabilities in rank histograms using large training datasets (Eckel and Walters, 1998). On the other hand, the "best member dressing" method represents a more complicated approach, aiming to scrutinize the members of an ensemble and select their best constituent, to identify the empirical distribution of the errors of that ensemble member, which is then used to "dress" each forecast (Roulston and Smith, 2003). Yet, contrary to Bayesian averaging, the best member dressing method is underlain by the premise that the best component can be identified with high probability, and as such does not take into account the uncertainty about the selection of the best member. In this regard, Raftery et al. (2003) presented an interesting approach to Bayesian model averaging using a finite mixture model coupled with the expectation-maximization algorithm (Dempster et al., 1977; McLachlan and Krishnan, 1997). The latter approach allows one to recast the problem of the optimal ensemble member selection in the form of latent or unobserved quantities and to subsequently maximize the likelihood function across a range of spatiotemporal domains (Raftery et al., 2003).

Given that the model calibration presented herein is effectively an inverse solution exercise, the examination of the posterior parameterization of the two models is an essential way to understand their ecological foundation and to put the respective predictions as well as their subsequent synthesis into perspective. First, we highlight the consistency between the two models with regards to the characterization of the generic phytoplankton group. In particular, both models are founded upon relatively high maximum phytoplankton growth rates (>2 day^{-1}), similar phytoplankton response to light availability (120–140 MJ m^{-2} day^{-1}), fast phosphorus kinetics (<10 μg P L^{-1}) and uptake rates (0.02 μg P L^{-1} day^{-1}), but differ on the relative magnitudes assigned to the phytoplankton biomass loss rates due to respiration and sinking. Namely, the complex model suggests lower metabolic rates (0.027 day^{-1}) and higher settling velocity (0.185 m day^{-1}), whereas the simple model places more emphasis on the former rate (0.09 day^{-1}) and downplays the latter one (0.05 m day^{-1}). Comparing these posterior characterizations with the parameterization of the three phytoplankton functional groups presented by Gudimov et al. (2010), we infer that our generic "phytoplankton" seemingly lies between the spring phytoplankton assemblage, primarily dominated (>80%) by diatoms (*Fragilaria crotonensis*, *Stephanodiscus niagarae*) and cryptophytes (*Rhodomonas minuta*, *Cryptomonas reflexa*), and the summer algal community, typically composed of dinophytes (*Gymnodinium helveticum*, *Ceratium furcoides*), chrysophytes (*Ochromonas sp*, *Dinobryon divergens*), and chlorophytes (*Coenochloris pyrenoidosa*, *Scenedesmus braziliensis*, *Coelastrum reticulatum*) (Munawar and Fitzpatrick, 2007). The updating of the two models with the calibration dataset resulted in similar zooplankton grazing (≈0.5 day^{-1}) and mortality rates (0.11–0.15 day^{-1}), intended to represent a zooplankton community typically dominated by cladocerans (*Bosmina longirostris*, *Daphnia*

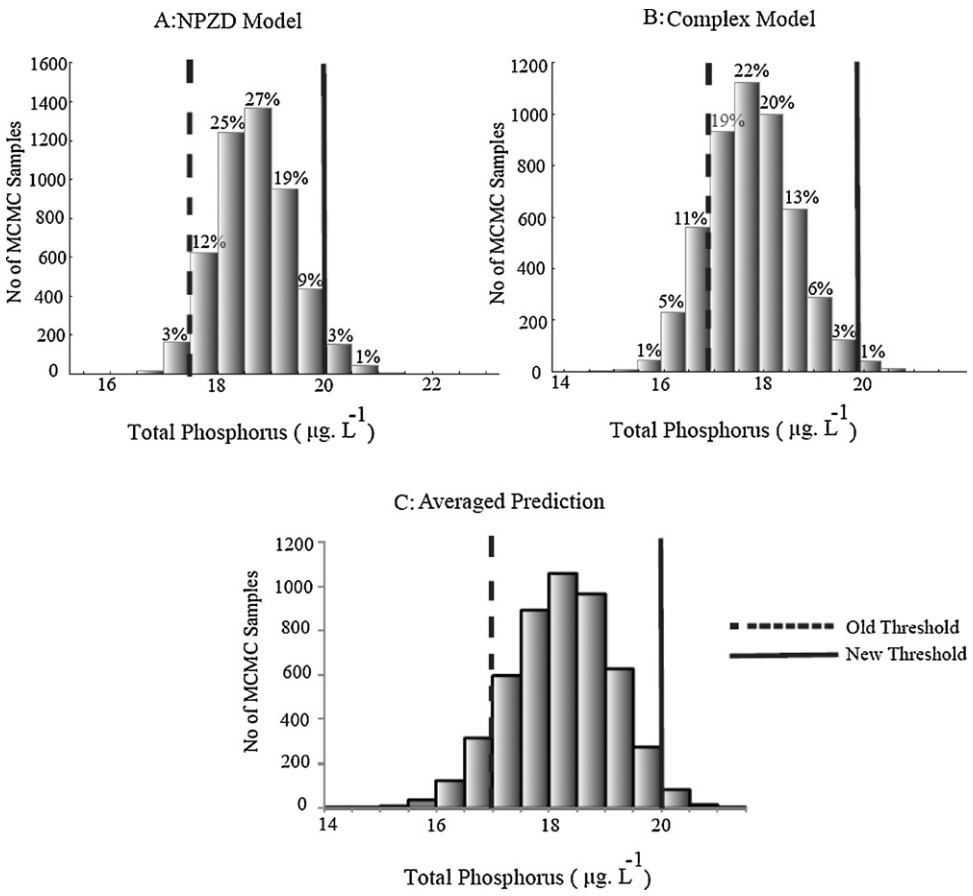


Fig. 7. Predictions of the epilimnetic summer total phosphorus concentrations, under the proposed nutrient loading reductions by the Hamilton Harbour RAP, based on the two eutrophication models (*A*–*B*) and their averaged predictions (*C*).

retrocurva, *Daphnia galeata mendotae*, *Ceriodaphnia lagustris*) and cyclopoid copepods (*Diacyclops thomasi*, *Cyclops vernalis*) relative to calanoids (*Leptodiaptomus* siciloides) (Gerlofsma et al., 2007). Finally, the posterior parameterization of the two models was also similar with respect to the sedimentation rate of particulate matter (>0.4 m day^{-1}), and the relative importance of the two factors that determine the illumination of the water column, i.e., the light extinction due to chlorophyll *a* (0.02–0.03 L (µg chl*a* L)$^{-1}$ m^{-1}), and the background light attenuation (≈0.2 m^{-1}).

Regarding the additional state variables of the complex model, we note that the prior cyanobacteria specification did not change after the calibration of the complex model, while the large process error value (σ_{CYAepi}) reflects the underestimation of its summer biomass levels (Fig. SI-2). The latter result suggests that whether the MCMC sampling was delimited in a suboptimal region of the posterior space or some of the assumptions (e.g., parameters not considered by the calibration exercise; Table SI-3) related to the present model structure disallow a better representation of the cyanobacteria patterns. In a similar manner, the marginal posteriors of the parameters associated with the nitrogen utilization by the two phytoplankton functional groups ($AH_{(cy)}/_{(phyt)}$, $NH_{(cy)}/_{(phyt)}$) remained unaltered with regard to their central tendency, dispersion, and shape. Although this result may partly support the soundness of the initial priors assigned, we believe that it primarily stems from the fact that the summer planktonic patterns are predominantly shaped by phosphate availability in the Harbour. On the other hand, while nitrate remains well above the growth-limiting levels, as reflected by the corresponding half saturation values, the low summer ammonium concentrations can presumably be a regulatory factor. Yet, the use of Liebig's Law of the Minimum in the algal growth term prohibits dissolved inorganic nitrogen from playing any role in the summer epilimnetic phytoplankton dynamics (see Ramin et al., 2011; Table 1 in their Electronic Supplementary Material). Interestingly, among the different strategies examined to improve the fit of the cyanobacteria-like group (i.e., different initialization of the two MCMC chains, alternative specification of the cyanobacteria parameter priors), we found that the most effective strategy is related to the values assigned to the strength of the ammonium inhibition for nitrate uptake (ψ); a parameter that was not part of our calibration exercise. The results presented herein were based on a ψ value equal to 0.05 for both species, which prioritizes ammonium uptake but does allow for nitrate utilization by our simulated phytoplankton community. When the nitrate uptake is practically switched off ($\psi > 0.1$) though, the relative abundance of our cyanobacteria-like group appears to be substantially promoted during the summer stratified period. However, we believe that the likelihood of this ecological mechanism actually to occur in the system is low, as it is extremely unlikely that the nitrate pool remains entirely unutilized. Further, this assumption contradicts Blomqvist et al. (1994) assertions, who argued that non-N-fixing cyanobacteria do have high competitiveness for ammonium but mainly in nitrate depleted systems, while eukaryotic phytoplankton develop when nitrate is the main N-component present. Thus, unless nitrogen is treated as an interactively essential resource (i.e., algal growth expressed as the product of two or more Monod terms), our current phytoplankton conceptualization is apparently inadequate to elucidate the conditions under which cyanobacteria outcompete other residents of the summer phytoplankton community and ultimately dominate the system.

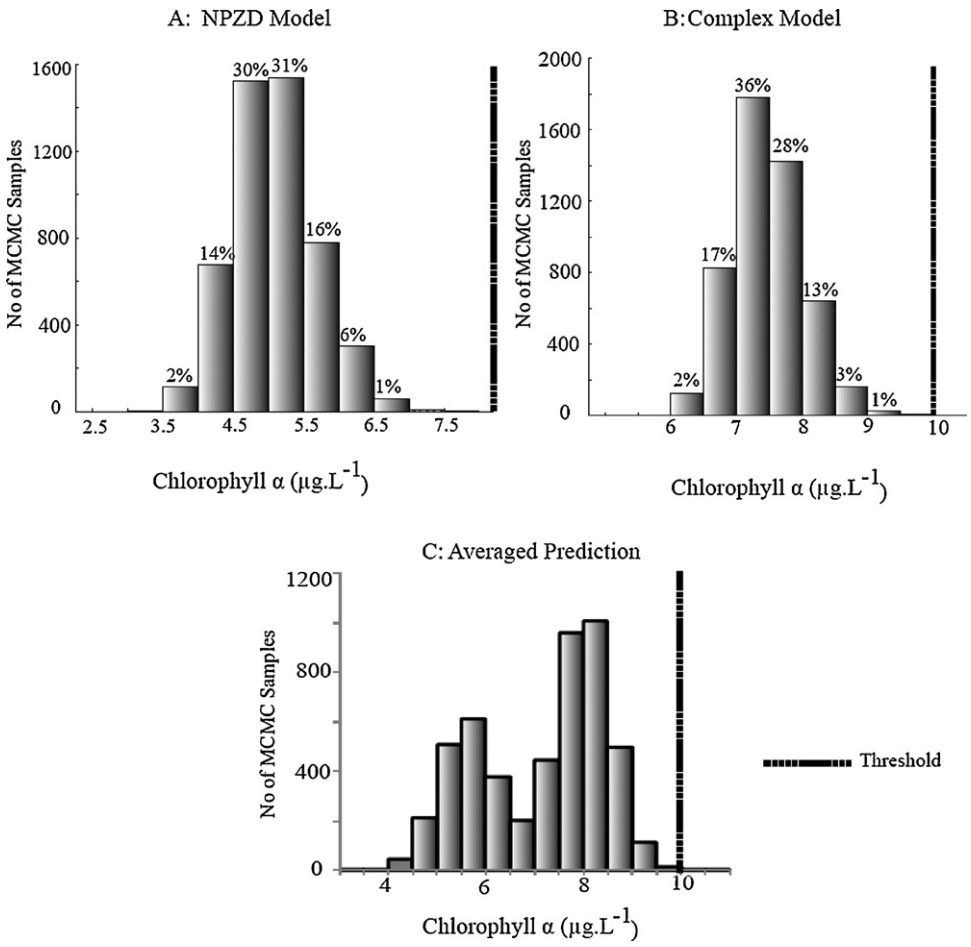


Fig. 8. Predictions of the epilimnetic summer chlorophyll *a* concentrations, under the proposed nutrient loading reductions by the Hamilton Harbour RAP, based on the two eutrophication models (*A–B*) and their averaged predictions (*C*).

Consistent with our earlier work (Gudimov et al., 2010; Ramin et al., 2011), both models downplay the role of particulate phosphorus mineralization (φ and $KP_{refmineral}$ <0.015 day^{-1}) as a source that can potentially replenish bioavailable phosphorus in the epilimnion. The allocation of a high proportion of plankton metabolism to directly replete the epilimnetic phosphate is evidently the most effective strategy to simultaneously match the typically high summer chl *a* levels (>15 μg L^{-1}) and low phosphate concentrations (<2–3 μg L^{-1}) in the Harbour epilimnion. Yet, contrary to our previous practice, it is interesting to note that here this pattern is derived through the calibration process instead of being postulated prior to the parameterization of the two models. According to the present analysis, a fraction between 30-60% of the plankton metabolism subsidizes the epilimnetic phosphate and may be responsible for a periodic disconnect between summer phytoplankton growth and exogenous nutrient loading variability. The question arising though is how realistic is it for such an active nutrient regeneration to occur in the present eutrophic state of the system and to what extent can this feedback loop modulate the phytoplankton response to the anticipated nutrient loading reductions? In principle, the rapid nutrient turnover rates can partly explain the relatively small fraction of dissolved inorganic phosphorus relative to the total phosphorus pool as well as the epilimnetic phytoplankton levels in the Harbour (Burley, 2007). The likelihood of an intense microbially mediated recycling, whereby nutrients are returned into the system in short time scales (<1 day) with minimal losses, has been repeatedly discussed in the literature over the last three decades (Goldman, 1984; Stone and Berman, 1993; Arhonditsis et al., 2004). Further, Gudimov et al. (2011) showed that regardless of the exogenous nutrient loading reductions, the fluxes from the sediments, and/or the control exerted from zooplankton, the exceedance frequency of the chlorophyll *a* target of 10 μg L^{-1} is consistently high when high nutrient recycling rates are assumed. Yet, before casting doubt on the anticipated efficacy of the on-going restoration efforts, the same study stressed that other potentially important nutrient sources, i.e., internal loading, episodic meteorological events (e.g., spring thaw, intense summer storms) and short-term variability at the local wastewater treatment plants, may also intermittently fuel epilimnetic algal growth. Thus, the relative contribution of the microbial loop as a nutrient supplier may be overstated (Gudimov et al., 2011).

The evaluation of performance of rival modelling strategies has received considerable attention in the literature, and the most popular proposition favours the distinction between training and test datasets (Hoeting et al., 1999). For probabilistic predictions, this strategy can potentially control both predictive bias (a systematic tendency to over- or underpredict the observed data), and lack of calibration (a systematic tendency to over- or understate predictive accuracy). By contrast, the practice followed herein involved the training of the two eutrophication models with a dataset that represents the water quality conditions typically prevailing in the Harbour, while no calibration or predictive validation of a specific time period was undertaken. As discussed in Gudimov et al. (2010; p. 527), this approach was selected as a pragmatic means to overcome the substantial uncertainty characterizing the exogenous nutrient loading estimates. Given the lack of reliable year-specific

nutrient loading inputs, a calibration or validation exercise framed upon year-to-year variations would have entailed the delineation of an ecological parameterization that links a highly uncertain forcing function to the ambient water quality conditions for a particular period. Rather, we adopted a more conservative strategy that uses models forced with the average (and thus more reliable) loading conditions to reproduce the average planktonic patterns in the system, and therefore avoid getting good fit to the observed data by introducing a series of errors that cancel each other out. It is also interesting to note that two recent papers extensively discussed the predictive capacity of the complex model (Gudimov et al., 2010; Ramin et al., 2011). Because of the uncertainty pertaining to the year-specific loading values from the early 90s, when the system was hyper-eutrophic, we opted for a somewhat different type of validation. Namely, rather than adopting any of the conventional types of predictive validation, we examined the ability of the model to reproduce the historical empirical relationships between annual P loading and summer *TP* and *chla*. The latter equations were originally used to guide the loading reductions in the system, and our exercise was pretty revealing about the credibility of some assumptions made to guide the contemporary management actions in the area.

However, Gudimov et al. (2011) noted that this strategy entails the risk of misrepresenting the actual range of system dynamics experienced when misleadingly phasing out short-term shifts of the year-to-year variability. In particular, a basic assumption of the Gudimov et al. (2010) and Ramin et al. (2011) calibration exercises was that the average summer epilimnetic *TP* revolved around the level of 30 $\mu g\,L^{-1}$, but recent monitoring evidence suggests that concentrations of 35–38 $\mu g\,L^{-1}$ are fairly typical at the offshore areas of the Harbour. Thus, one of the objectives of the present analysis was to examine the likelihood of the system to still meet the originally proposed *TP* water quality goal of 17 $\mu g\,L^{-1}$, if the starting point is higher by 5–8 μg *TP* L^{-1} relative to the reference conditions used from the earlier modelling studies. Both models predict that the current epilimnetic total phosphorus criterion of 17 $\mu g\,L^{-1}$ is stringent if the current summer epilimnetic *TP* concentrations are set to an average level of 35–38 $\mu g\,L^{-1}$. On a positive note though, and consistent with Gudimov et al.'s (2011) predictions, our analysis also shows that a somewhat higher value (e.g., 20 $\mu g\,L^{-1}$) may provide a more pragmatic goal, although the uncertainty surrounding the exogenous nutrient loading still makes it compelling to adopt water quality criteria that permit a realistic frequency level of violations.

Our modelling analysis also suggests that a target of mean chlorophyll *a* levels in the Harbour lower than 10 $\mu g\,L^{-1}$ is achievable, as both models predict that the average epilimnetic concentrations will fall below this threshold value. Yet, we also found that the simple model provides an optimistic prediction that the average summer chlorophyll *a* may even reach the level of 5 μg *chla* L^{-1}. Despite the underestimation of the summer phytoplankton biomass, the complex model clearly supports more conservative predictions ($\approx$7.5 μg *chla* L^{-1}), and thus the question arising is which factor may be driving the optimism of the simple model and to what extent this statement is realistic? One of the major structural differences of the two models lies in the way they handle the nutrient fluxes from the sediments, i.e., a static phosphorus flux vis-à-vis a mechanistic characterization that relates phosphorus release to particulate sedimentation and burial rates (Ramin et al., 2011). Being part of the model updating process, the simple model predicts that the sediments contribute approximately 1.1 mg *P* m^2 day^{-1} into the overlying water column, whereas the same fluxes are raised up to 2.0 mg *P* m^2 day^{-1} with the complex model. Notably, Azcue et al. (1998) reported upward diffusion PO_4 fluxes in the Harbour at the level of 1.7 mg m^2 day^{-1}, which appear to be somewhat closer to the latter estimate. Under the reduced nutrient loading scenario, the dynamic nature of the sediment response with the complex model decreases the release of phosphorus at the level of 1.5 mg m^2 day^{-1}, which however remains well above the flux used to force the simple model. Because the simple model also fails to capture the summer hypolimnetic phosphate accumulation, we believe that this discrepancy most likely reflects one of its structural weaknesses and also highlights the importance of embracing more sophisticated approaches to sediment diagenesis in the Harbour (Dittrich et al., 2009; Trolle et al., 2010). Despite all the arguments historically used to downplay the relative contribution of the sediment fluxes in the system, recent evidence suggests that the hypolimnetic phosphate can easily exceed the level of 30 μg PO_4 L^{-1} for extended period (3–4 weeks) during the late summer/early fall (T. Labencki, unpublished data). This pattern likely suggests that the summer epilimnetic environment may also be subjected to intermittent nutrient pulses from the hypolimnion, which in turn can have profound ramifications on the dynamics of the phytoplankton community (Jorgensen and Padisak, 1996; Soranno, 1997).

The simplified spatial segmentation of both eutrophication models and the inability to account for persistent spatial gradients or other hot spots in the system is another source of uncertainty that can potentially shape the probabilistic statements made by our modelling analyses. Rao et al. (2009) showed that the mean summer circulation pattern in the Harbour was characterized by two major counter-rotating gyres that play significant role in transporting contaminants within the system and thus explaining their distribution patterns in its north–eastern sector. Likewise, Burley (2007) reported distinct differences in the nutrient concentrations and chlorophyll *a* levels between inshore and offshore sites in the Harbour. The high computational demands of a spatially explicit model, however, may present an obstacle to the efficient application of Bayesian calibration. Recent attempts to improve the computational efficiency of MCMC implementations of Bayesian inference for water quality models have focused on the development of parallel algorithms (Whiley and Wilson, 2004; Altekar et al., 2004). Parallel computation for MCMC can reduce the time needed to generate a sufficient number of samples from target distributions of larger dimensions, although Whiley and Wilson (2004) assert that a good proposal distribution is of equal importance as the implementation of a parallelization scheme. Other propositions to efficiently estimate the posterior probability density function of parameters in complex high-dimensional problems involve the development of adaptive MCMC schemes that ensure ergodicity while adjusting the scale and orientation of the proposal distributions, e.g., the differential evolution adaptive Metropolis (DREAM) introduced by Vrugt et al. (2008). Another promising prospect involves the construction of one-dimensional frameworks in which various complexity model structures and physical forcing regimes could be implemented and calibrated using rigorous parameter estimation techniques (Friedrichs et al., 2007). For example, Schartau and Oschlies (2003) presented such an approach to calibrate a simple biogeochemical model simultaneously at three sites using a genetic algorithm and physical forcing from a 3-D hydrodynamic model. By doing so, much of the hydrodynamic computations are performed externally and along with the simplified spatial structure of the ecosystem model, the computational demands are significantly reduced. We believe that such 1-D Bayesian parameterization in three or more sites based on 3-D hydrodynamic forcing will be sufficient to accommodate the spatial variability in the Hamilton Harbour.

In conclusion, we used two models of different complexity to examine the robustness of predictive statements made from our earlier work regarding the likelihood of the Hamilton Harbour Area of Concern to meet the delisting objectives for the beneficial use impairment (*BUI*) *Eutrophication or Undesirable Algae*, if the nutrient

loading reductions proposed by the Hamilton Harbour Remedial Action Plan are actually implemented. The structural differences between the two models were primarily intended to: (*i*) evaluate the implications of alternative quantification strategies of nutrient release rates from the sediments, as we project future system responses, and (*ii*) examine our capacity to realistically depict the summer phytoplankton community composition under the present loading conditions. The basic lessons learned from the present analysis are as follows:

- The current epilimnetic total phosphorus goal of 17 $\mu g\,L^{-1}$ is probably too stringent and therefore a somewhat higher mean value (e.g., 20 $\mu g\,L^{-1}$) may provide a more realistic target, when the starting point for our projections is higher by 5–8 $\mu g\ TP\ L^{-1}$, i.e., a summer average of 35–38 $\mu g\ TP\ L^{-1}$ instead of 30 $\mu g\ TP\ L^{-1}$. In a strict numerical sense, the selection of the training dataset to parameterize a mathematical model is particularly critical in determining the impact of environmental management plans. If we strictly adhere to a single numerical value, our assumptions on what should be perceived as typical water quality conditions in the Harbour can evidently make a difference on the conclusions drawn about the most likely system trajectory. Acknowledging also the uncertainty of the contemporary nutrient loading estimates in the Harbour as well as the lurking known or unknown "ecological unknowns" (sensu Gudimov et al., 2011), we caution that the water quality setting process must be pragmatic and the natural variability should be explicitly accommodated by permitting a realistic frequency of violations, e.g., exceedences of the goal for 10–15% or less of the weekly-collected samples during the stratified period should still be considered as system compliance. It is important to remember that the water quality criteria are merely proxies of the desirable water quality conditions and therefore other quantitative or even qualitative features/indices of the ecosystem functioning may be equally insightful to determine what is successful restoration.
- The target of mean chlorophyll *a* concentrations in the Harbour lower than 10 $\mu g\,L^{-1}$ is achievable. Yet, we note that the complex model appears to underestimate the summer epilimnetic phytoplankton biomass, whereas the simple one supports an over-optimistic prediction that the chlorophyll *a* levels can ultimately reach the low level of 5 $\mu g\,L^{-1}$.
- The latter discrepancy most likely stems from the simplified approach adopted to account for the role of the sediments with the simple model. Based also on our earlier work (Ramin et al., 2011; Gudimov et al., 2011), the present analysis reiterates the importance of revisiting our current perception about the role of internal loading in the Harbour.
- Our limited ability to effectively reproduce the phytoplankton seasonal succession patterns and the gradual cyanobacteria biomass increase towards the end of the summer pinpoints one of the knowledge gaps and outstanding challenges of the on-going restoration efforts. Following Watson et al's (2008) standpoint, we advocate that future research should aim to integrate aspects of all the single-factor hypotheses presented in the literature, such as the buoyancy regulation, ability to fix molecular nitrogen, *N*–*P* ratios, minimization of mortality through an immunity to grazing by zooplankton, ability to out-compete most other phytoplankton for ammonium-nitrogen, and elevated iron levels in the system. It becomes increasingly clear that there are more than one or two causal factors underlying the patterns of cyanobacteria dominance and their capacity to outcompete the usual eukaryotic residents of the summer phytoplankton communities (e.g., chlorophytes).
- Modellers must acknowledge the uncertainty pertaining to the selection of the optimal model structure for a specific environmental management problem, and Bayesian averaging of two or more models is a promising means for improving the contemporary modelling practice. In the context of ecological process-based modelling though, this approach should not be viewed solely as a framework to improve our predictive devices, but rather as an opportunity to compare alternative ecological structures, to challenge existing ecosystem conceptualizations, and to integrate across different (and often conflicting) paradigms. Future research should also focus on the refinement of the weighting schemes and other performance standards to impartially synthesize the predictions of different models. Several interesting statistical post-processing methods presented in the field of ensemble weather forecasting will greatly benefit our attempts to develop weighting schemes suitable for the synthesis of multiple ecosystem models. Some of the outstanding challenges involve the development of ground rules for the features of the calibration and validation domain (Reckhow and Chapra, 1999; Raftery et al., 2003; Anderson, 2005), the inclusion of penalties for model complexity that will allow building forecasts upon parsimonious models (McDonald and Urban, 2010), and performance assessment that does not exclusively consider model endpoints but also examines the plausibility of the underlying ecosystem structures, i.e., biological rates, ecological processes or derived quantities (Arhonditsis and Brett, 2004)

Acknowledgments

This project has received funding support from the Ontario Ministry of the Environment (Canada-Ontario Grant Agreement 120808). Such support does not indicate endorsement by the Ministry of the contents of this material. Maryam Ramin has received support from the Natural Sciences and Engineering Research Council of Canada (Doctoral Scholarships). All the material pertinent to this analysis is available upon request from the corresponding author.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ecolmodel.2012.05.023.

References

Altekar, G., Dwarkadas, S., Huelsenbeck, J.P., Ronquist, F., 2004. Parallel Metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. Bioinformatics 20, 407–415.

Anderson, T.R., 2005. Plankton functional type modelling: running before we can walk? Journal of Plankton Research 27, 1073–1081.

Arhonditsis, G., Eleftheriadou, M., Karydis, M., Tsirtsis, G., 2003. Eutrophication risk assessment in coastal embayments using simple statistical models. Marine Pollution Bulletin 46, 1174–1178.

Arhonditsis, G.B., Brett, M.T., 2004. Evaluation of the current state of mechanistic aquatic biogeochemical modeling. Marine Ecology: Progress Series 271, 13–26.

Arhonditsis, G.B., Winder, M., Brett, M.T., Schindler, D.E., 2004. Patterns and mechanisms of phytoplankton variability in Lake Washington (USA). Water Research 38, 4013–4027.

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

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

Arhonditsis, G.B., Qian, S.S., Stow, C.A., Lamon, E.C., Reckhow, K.H., 2007. Eutrophication risk assessment using Bayesian calibration of process-based models: application to a mesotrophic lake. Ecological Modelling 208, 215–229.

Arhonditsis, G.B., Papantou, D., Zhang, W.T., Perhar, G., Massos, E., Shi, M.L., 2008a. Bayesian calibration of mechanistic aquatic biogeochemical models and benefits for environmental management. Journal of Marine Systems 73, 8–30.

Arhonditsis, G.B., Perhar, G., Zhang, W.T., Massos, E., Shi, M., Das, A., 2008b. Addressing equifinality and uncertainty in eutrophication models. Water Resources Research 44, W01420.

Arhonditsis, G.B., 2009. Useless arithmetic? Lessons learnt from aquatic biogeochemical modeling. In: Hanrahan, G. (Ed.), Modelling of Pollutants in Complex Environmental Systems. ILM Publications, pp. 3–26.
Azcue, J.M., Zeman, A.J., Mudroch, A., Rosa, F., Patterson, T., 1998. Assessment of sediment Harbour, Canada. Water Science and Technology 37, 323–329.
Azim, M.A., Kumarappah, A., Bhavsar, S.P., Backus, S.M., Arhonditsis, G.B., 2011. Detection of the spatiotemporal trends of mercury in Lake Erie Fish Communities: a Bayesian approach. Environmental Science and Technology 45, 2217–2226.
Bao, L., Gneiting, T., Grimit, E.P., Guttorp, P., Raftery, A.D., 2010. Bias correction and Bayesian model averaging for ensemble forecasts of surface wind direction. Monthly Weather Review 138, 1811–1821.
Berounsky, V.M., Nixon, S.W., 1990. Temperature and the annual cycle of nitrification in waters of Narragansett Bay. Limnology and Oceanography 35, 1610–1617.
Beven, K., Freer, J., 2001. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. Journal of Hydrology 249, 11–29.
Blomqvist, P., Pettersson, A., Hyenstrand, P., 1994. Ammonium–nitrogen—a key regulatory factor causing dominance of non-nitrogen fixing cyanobacteria in aquatic systems. Archiv Fur Hydrobiologie 132, 141–164.
Borsuk, M.E., Stow, C.A., Reckhow, K.H., 2004. A Bayesian network of eutrophication models for synthesis, prediction, and uncertainty analysis. Ecological Modelling 173, 219–239.
Brooks, S.P., Gelman, A., 1998. Alternative methods for monitoring convergence of iterative simulations. Journal of Computational and Graphical Statistics 7, 434–455.
Burley, M., 2007. Water quality and phytoplankton photosynthesis. Canadian Technical Report on Fisheries and Aquatic Sciences 2729, 9–42.
Cerco, C., Cole, T., 1993. 3-Dimensional eutrophication model of Chesapeake Bay. Journal of Environmental Engineering: ASCE 119, 1006–1025.
Chen, C.S., Ji, R.B., Schwab, D.J., Beletsky, D., Fahnenstiel, G.L., Jiang, M.S., Johengen, T.H., Vanderploeg, H., Eadie, B., Budd, J.W., Bundy, M.H., Gardner, W., Cotner, J., Lavrentyev, P.J., 2002. A model study of the coupled biological and physical dynamics in Lake Michigan. Ecological Modelling 152, 145–168.
Christakos, G., 2002. On the assimilation of uncertain physical knowledge bases Bayesian and non-Bayesian techniques. Advances in Water Resources 25, 1257–1274.
Christakos, G., 2003. Critical conceptualism in environmental modelling and prediction. Environmental Science and Technology 37, 4685–4693.
Clark, J.S., Carpenter, S.R., Barber, M., Collins, S., Dobson, A., Foley, J., Lodge, D.M., Pascual, M., Pielke Jr., R., Pizer, W., Pringle, C., Reid, W.V., Rose, K.A., Sala, O., Schlesinger, W.H., Wall, D.H., Wear, D., 2001. Ecological forecasts: an emerging imperative. Science 293, 657–660.
Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the *EM* algorithm. Journal of Royal Statistical Society Series B 34, 1–38.
Dermott, R., Johannsson, O., Munawar, M., Bonnell, R., Bowen, K., Burley, M., Fitzpatrick, M., Gerlofsma, J., Niblock, H., 2007. Assessment of lower food web in Hamilton Harbour Lake Ontario, 2002–2004. Canadian Technical Report on Fisheries and Aquatic Sciences 2729, 120.
Dittrich, M., Wehrli, B., Reichert, P., 2009. Lake sediments during the transient eutrophication period: reactive-transport model and identifiability study. Ecological Modelling 220, 2751–2769.
Eckel, F., Walters, M., 1998. Calibrated probabilistic quantitative precipitation forecasts based on the MRF ensemble. Weather Forecast 13, 1132–1147.
Edwards, A.M., 2001. Adding detritus to a nutrient-phytoplankton–zooplankton model: a dynamical-systems approach. Journal of Plankton Research 23, 389–413.
Endres, D.M., Schindelin, J.E., 2003. A new metric for probability distributions. IEEE Transactions on Information Theory 49, 1858–1860.
Flynn, K.J., 2005. Castles built on sand: dysfunctionality in plankton models and the inadequacy of dialogue between biologists and modellers. Journal of Plankton Research 27, 1205–1210.
Franks, P.J.S., 1995. Coupled physical–biological models in oceanography. Reviews of Geophysics 33, 1177–1187.
Friedrichs, M.A.M., Dusenberry, J.A., Anderson, L.A., Armstrong, R.A., Chai, F., Christian, J.R., Doney, S.C., Dunne, J., Fujii, M., Hood, R., McGillicuddy Jr., D.J., Moore, J.K., Schartau, M., Spitz, Y.H., Wiggert, J.D., 2007. Assessment of skill and portability in regional marine biogeochemical models: role of multiple planktonic groups. Journal of Geophysical Research 112, http://dx.doi.org/10.1029/2006JC003852.
Gauch, H.G., 1993. Prediction, parsimony and noise. American Scientist 81, 468–478.
Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 1995. Bayesian Data Analysis. Chapman and Hall, New York.
Gerlofsma, J., Bowen, K., Johannsson, O., 2007. Zooplankton in Hamilton Harbour 2002–2004. Canadian Technical Report on Fisheries and Aquatic Sciences 2729, 65–90.
Gilks, W.R., Richardson, S., Spiegelhalter, D.J., 1998. Markov Chain Monte Carlo in Practice. Chapman & Hall/CRC, 512 p.
Goldman, J.C., 1984. Oceanic nutrient cycles. In: Fasham, M.J. (Ed.), Flows of Energy and Materials in Marine Ecosystems: Theory and Practice. Plenum Press, London, pp. 137–170.
Gudimov, A., Stremilov, S., Ramin, M., Arhonditsis, G.B., 2010. Eutrophication risk assessment in Hamilton Harbour System analysis and evaluation of nutrient loading scenarios. Journal of Great Lakes Research 36, 520–539.
Gudimov, A., Ramin, M., Labencki, T., Wellen, C., Shelar, M., Shimoda1, Y., Boyd, D., Arhonditsis, G.B., 2011. Predicting the response of Hamilton Harbour to the nutrient loading reductions: a modeling analysis of the ecological unknowns. Journal of Great Lakes Research 37, 494–506.
Hamilton Harbour RAP Stakeholder Forum, 2003. Remedial Action Plan for Hamilton Harbour: Stage 2 Update 2002. Burlington, Ontario.
Hamilton, D.P., Schladow, S.G., 1997. Prediction of water quality in lakes and reservoirs. Part I—Model description. Ecological Modelling 96, 91–110.
Hiriart-Baer, V.P., Milne, J., Charlton, M.N., 2009. Water quality trends in Hamilton Harbour: two decades of change in nutrients and chlorophyll a. Journal of Great Lakes Research 35, 293–301.
Hoeting, J.A., Madigan, D., Raftery, A.E., Volinsky, C.T., 1999. Bayesian model averaging: a tutorial. Statistical Science 14, 382–417.
Hong, B.G., Strawderman, R.L., Swaney, D.P., Weinstein, D.A., 2005. Bayesian estimation of input parameters of a nitrogen cycle model applied to a forested reference watershed, Hubbard Brook Watershed Six. Water Resources Research 41, W03007, http://dx.doi.org/10.1029/2004WR003551.
Jorgensen, S.E., Nielsen, S.N., Jorgensen, L.A., 1991. Handbook of Ecological Parameters and Ecotoxicology. Pergamon Press, Amsterdam.
Jorgensen, S.E., Padisak, J., 1996. Does the intermediate disturbance hypothesis comply with thermodynamics? Hydrobiologia 323, 9–21.
Klapwijk, A., Snodgrass, W.J., 1985. Model for Lake-Bay exchange flow. Journal of Great Lakes Research 11, 43–52.
Kohli, B., 1979. Mass exchange between Hamilton Harbour and Lake Ontario. Journal of Great Lakes Research 5, 36–44.
Lamon, E.C., Clyde, M., 2000. Accounting for model uncertainty in prediction of chlorophyll *a* in Lake Okeechobee. Journal of Agricultural Biological and Environmental Statistics 5, 297–322.
Lampert, W., Sommer, U., 1997. Limnoecology: The Ecology of Lakes and Streams. Oxford University Press.
Law, T., Zhang, W., Zhao, J., Arhonditsis, G.B., 2009. Structural changes in lake functioning induced from nutrient loading and climate variability. Ecological Modelling 220, 979–997.
McLachlan, G., Krishnan, T., 1997. The EM Algorithm and Extensions. Wiley Series in Probability and Statistics, John Wiley & Sons.
McDonald, C., Urban, N., 2010. Using a model selection criterion to identify appropriate complexity in aquatic biogeochemical models. Ecological Modelling 221, 428–432.
Munawar, M., Fitzpatrick, M., 2007. An integrated assessment of the microbial and planktonic communities of Hamilton Harbour. Canadian Technical Report on Fisheries and Aquatic Sciences 2729, 43–63.
Neal, R., 1998. Suppressing random walks in Markov chain Monte Carlo using ordered over-relaxation. In: Jordan, M.I. (Ed.), Learning in Graphical Models. Kluwer Academic Publishers, Dordrecht, pp. 205–230.
Neuman, W.L., 2003. Maximum likelihood Bayesian averaging of uncertain model predictions. Stochastic Environmental Research and Risk Assessment 17, 291–305.
Omlin, M., Brun, R., Reichert, P., 2001. Biogeochemical model of Lake Zurich: sensitivity, identifiability and uncertainty analysis. Ecological Modelling 141 (1-3), 105–123.
Oreskes, N., Shrader-Frechette, K., Belitz, K., 1994. Verification validation, and confirmation of numerical models in the earth sciences. Science 263, 641–646.
Peters, R.H., 1986. The role of prediction in limnology. Limnology and Oceanography 31, 1143–1159.
Pilkey, O.H., Pilkey-Jarvis, L., 2007. Useless Arithmetic: Why Environmental Scientists Can't Predict the Future. Columbia University Press, New York, New York, p. 230.
Qian, S.S., Reckhow, K.H., 2007. Combining model results and monitoring data for water quality assessment. Environmental Science and Technology 41, 5008–5013.
Raftery, A.E., Balabdaoui, F., Gneiting, T., Polakowski, M., 2003. Using Bayesian model averaging to calibrate forecast ensembles, Technical Report 440, Department of Statistics, University of Washington, Seattle.
Raftery, A.E., Gneiting, T., Balabdaoui, F., Polakowski, M., 2005. Using Bayesian model averaging to calibrate forecast ensembles. Monthly Weather Review 133, 1155–1174.
Ramin, M., Stremilov, S., Labencki, T., Gudimov, A., Boyd, D., Arhonditsis, G.B., 2011. Integration of mathematical modeling and Bayesian inference for setting water quality criteria in Hamilton Harbour, Ontario, Canada. Environmental Modelling & Software 26, 337–353.
Rao, Y.R., Marvin, C.H., Zhao, J., 2009. Application of a numerical model for circulation, temperature and pollutant distribution in Hamilton Harbour. Journal of Great Lakes Research 35, 61–73.
Reckhow, K.H., Chapra, S.C., 1999. Modeling excessive nutrient loading in the environment. Environmental Pollution 100, 197–207.
Reichert, P., Omlin, M., 1997. On the usefulness of overparameterized ecological models. Ecological Modelling 95, 289–299.
Reynolds, C.S., 1984. The Ecology of Freshwater Phytoplankton. Cambridge University Press, Cambridge, UK.
Reynolds, C.S., 2006. The Ecology of Phytoplankton. Cambridge University Press.
Roulston, M.S., Smith, L.A., 2003. Combining dynamical and statistical ensembles. Tellus Series A 55, 16–30.
Rykiel Jr., E.J., 1996. Testing ecological models: the meaning of validation. Ecological Modelling 90, 229–244.
Sandgren, C.D., 1991. Growth and Reproductive Strategies of Freshwater Phytoplankton. Cambridge University Press.
Scavia, D., Liu, Y., 2009. Exploring estuarine nutrient susceptibility. Environmental Science and Technology 43, 3474–3479.

Schartau, M., Oschlies, A., 2003. Simultaneous data-based optimization of a 1D-ecosystem model at three locations in the North Atlantic Ocean. Part II—Standing stocks and nitrogen fluxes. Journal of Marine Research 61, 795–821.
Sloughter, J.M., Raftery, A.E., Gneiting, T., Fraley, C., 2007. Probabilistic quantitative precipitation forecasting using Bayesian model averaging. Monthly Weather Review 135, 3209–3220.
Sloughter, J.M., Gneiting, T., Raftery, A.E., 2010. Probabilistic wind speed forecasting using ensembles and Bayesian model averaging. Journal of the American Statistical Association 105, 25–35.
Sommer, U., 1989. Phytoplankton Ecology. Succession in Plankton Communities. Springer-Verlag.
Soranno, P.A., 1997. Factors affecting the timing of surface scums and epilimnetic blooms of blue-green algae in a eutrophic lake. Canadian Journal of Fisheries and Aquatic Sciences 54, 1965–1975.
Steinberg, L.J., Reckhow, K.H., Wolpert, R.L., 1997. Characterization of parameters in mechanistic models: a case study of a PCB fate and transport model. Ecological Modelling 97, 35–46.
Stone, L., Berman, T., 1993. Positive feedback in aquatic ecosystems. The case of the microbial loop. Bulletin of Mathematical Biology 55, 919–936.
Stow, C.A., Lamon, E.C., Qian, S.S., Schrank, C.S., 2004. Will Lake Michigan lake Trout meet the Great Lakes strategy 2002 reduction goal? Environmental Science and Technology 38, 359–363.
Trolle, D., Hamilton, D.P., Pilditch, C.A., 2010. Evaluating the influence of lake morphology, trophic status and diagenesis on geochemical profiles in lake sediments. Applied Geochemistry 25, 621–632.
Vrugt, J.A., Ter Braak, C.J.F., Clark, M.P., Hyman, J.M., Robinson, B., 2008. Treatment of input uncertainty in hydrologic modeling: doing hydrology backward with Markov chain Monte Carlo simulation. Water Resources Research 44, W00B09, http://dx.doi.org/10.1029/2007WR006720.
Watson, S.B., Ridal, J., Boyer, G.L., 2008. Taste and odour and cyanobacterial toxins: impairment, prediction, and management in the Great Lakes. Canadian Journal of Fisheries and Aquatic Sciences 65, 1779–1796.
Wetzel, R.G., 2001. Limnology: Lake and River Ecosystems, 3rd ed. Academic Press, New York, USA.
Whiley, M., Wilson, S.P., 2004. Parallel algorithms for Markov chain Monte Carlo methods in latent spatial Gaussian models. Journal of Statistical Computation 14, 171–179.
Wilks, D.S., 2002. Smoothing forecast ensembles with fitted probability distributions. Quarterly Journal of the Royal Meteorological Society 128, 2821–2836.
Zhang, W., Arhonditsis, G.B., 2008. Predicting the frequency of water quality standard violations using Bayesian calibration of eutrophication models. Journal of Great Lakes Research 34, 698–720.

A BAYESIAN SYNTHESIS OF PREDICTIONS FROM DIFFERENT MODELS FOR SETTING WATER QUALITY CRITERIA

(SUPPORTING INFORMATION)

Maryam Ramin[1], Tanya Labencki[2], Duncan Boyd[2], Dennis Trolle[3], George B. Arhonditsis[1*]

[1] Ecological Modelling Laboratory,

Department of Physical & Environmental Sciences, University of Toronto,

Toronto, Ontario, Canada, M1C 1A4

[2] Great Lakes Unit, Water Monitoring & Reporting Section, Ontario Ministry of the Environment, Environmental Monitoring and Reporting Branch, Toronto, Ontario, Canada, M9P 3V6

[3] Department of Freshwater Ecology, National Environmental Research Institute, Aarhus University, PO Box 314, 8600 Silkeborg, Denmark

* Corresponding author

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

1. Model description

Detailed description of the two models used herein can be found elsewhere (Arhonditsis *et al.*, 2007; Law *et al.*, 2009; Ramin *et al.*, 2011), and thus this section solely provides their basic conceptual design with emphasis on the distinct features from our earlier work.

1.1. Complex eutrophication model: The complex eutrophication model considers the interactions among the eight state variables: nitrate (NO_3), ammonium (NH_4), phosphate (PO_4), generic phytoplankton, cyanobacteria-like phytoplankton, zooplankton, organic nitrogen (*ON*) and organic phosphorus (*OP*). We considered a two-compartment vertical segmentation representing the epilimnion and hypolimnion of the Harbour. The depths of the two boxes varied with time and were explicitly defined based on extensive field measurements for the study period 1987–2007 (Dermott *et al.*, 2007; Hiriart-Baer *et al.*, 2009). The two phytoplankton functional groups simulated differ with regards to their strategies for resource competition (nitrogen, phosphorus, light, and temperature) and metabolic rates as well as their morphological features (settling velocities, self-shading effects). The cyanobacteria-like group is modeled as K-strategist with low maximum growth and metabolic rates, slow P and fast N kinetics, higher tolerance to low light availability, low settling velocities, and high temperature optima. By contrast, the more generic "phytoplankton" group aims to represent the rest of the phytoplankton community, in that it has attributes of r-selected organisms with high maximum growth rates and higher metabolic losses, strong phosphorus and weak nitrogen competition ability, lower tolerance to low light availability, low temperature optima, and high sinking velocities.

The governing equation for phytoplankton biomass accounts for phytoplankton production and losses due to mortality, settling, dreissenid filtration, and herbivorous zooplankton grazing. Phytoplankton growth is limited from the water temperature conditions, the nutrient and light availability. Phosphorus dynamics within the phytoplankton cells account for luxury uptake, i.e., phytoplankton nutrient uptake depends on both internal and external concentrations and is confined by upper and lower internal storage capacity (Arhonditsis *et al.*, 2002, Zhao *et al.*, 2008). Our model

explicitly considers the role of new and regenerated production using separate formulations that relate phytoplankton uptake to ambient nitrate and ammonium concentrations (Eppley–Peterson *f*-ratio paradigm; Eppley and Peterson, 1979). Temperature-modulated zooplankton grazing was modeled using a Michaelis–Menten equation and the assimilated fraction of the grazed material fuels growth. Zooplankton has three alternative food sources (the two phytoplankton groups and biogenic particulate matter or detritus) grazed with preference that changes dynamically as a function of their relative abundance alongside with a selective zooplankton preference for phytoplankton and detritus over cyanobacteria (Fasham *et al.*, 1990).

There are three nitrogen forms considered in the model: nitrate (NO_3), ammonium (NH_4), and organic nitrogen (*ON*). The ammonium equation considers the phytoplankton uptake and the proportion of phytoplankton and zooplankton mortality that is returned back to the system as ammonium ions. Ammonium is also oxidized to nitrate through nitrification and the kinetics of this process is modeled as a function of the ammonium, and the externally-forced dissolved oxygen, temperature, and light availability (Tian *et al.*, 2001). The nitrate equation also takes into account the amount of ammonium oxidized to nitrate through nitrification and the amount of nitrate lost as nitrogen gas through denitrification. The latter process is modeled as a function of dissolved oxygen, temperature and the contemporary nitrate concentrations (Arhonditsis and Brett, 2005). The organic nitrogen equation considers the contribution of phytoplankton and zooplankton mortality to the organic nitrogen pool and the temperature-dependent mineralization that transforms organic nitrogen to ammonium. Two state variables of the phosphorus cycle are considered in the model: phosphate (PO_4) and organic phosphorus (*OP*). The phosphate equation considers the phytoplankton uptake, the proportion of phytoplankton and zooplankton mortality/higher predation that is directly supplied into the system in inorganic form, the bacteria-mediated mineralization of organic phosphorus, and the net diffusive fluxes between adjacent compartments. The organic phosphorus equation also considers the amount of organic phosphorus that is redistributed through plankton basal metabolism. A fraction of organic phosphorus settles to the sediment and another fraction is mineralized to phosphate through a first-order reaction. We also

consider external nutrient loads to the system and losses via the exchanges with Lake Ontario (Gudimov *et al.*, 2010, 2011; Ramin *et al.*, 2011). A simple mechanistic approach was used to relate the fluxes of nitrogen and phosphorus from the sediments with the algal and particulate matter sedimentation and burial rates, while also accounting for the role of temperature and dissolved oxygen (Arhonditsis and Brett, 2005). The relative magnitudes of ammonium and nitrate fluxes were also determined by nitrification and denitrification occurring at the sediment surface.

1.2. Simple eutrophication model: We developed a simple model that considers the interplay among the limiting nutrient (phosphate), phytoplankton, zooplankton, and detritus (particulate phosphorus); also known as NPZD model in the literature. The spatial segmentation of the model consists of three compartments representing the epilimnion, thermocline, and hypolimnion of the system. The equation for phytoplankton biomass accounts for phytoplankton production, losses due to basal metabolism, herbivorous zooplankton grazing, and settling. Phytoplankton growth is directly linked to the ambient phosphorus concentrations without explicit consideration of the control exerted by the intra-cellular storage strategies. Phytoplankton basal metabolic losses include all internal processes that decrease algal biomass as well as natural mortality. The zooplankton biomass equation considers zooplankton growth and losses due to natural mortality and predation. Zooplankton feeds upon phytoplankton and detritus with kinetics described by the Holling Type III function. Contrary to our earlier work (Law *et al.*, 2009), the palatability of the two food sources (ω) is treated as a stochastic node assigned a prior distribution and subjected to updating by the calibration dataset. A fraction of the zooplankton grazing is assimilated and fuels growth, another fraction is excreted as phosphate, while the remaining fraction represents the faecal pellets contributing to the detritus pool. We assumed a unimodal response of the planktonic processes on temperature seasonal variability modeled by a Gaussian-like probability curve (Arhonditsis and Brett, 2005). The phosphate equation considers the phytoplankton uptake, the gains due to zooplankton excretion/predation, the bacteria-mediated mineralization of detritus, and the net diffusive fluxes between adjacent compartments. The detritus equation takes into account the contributions from phytoplankton respiration and zooplankton

excretion, and the losses due to bacteria-mediated mineralization and settling. The effects of the seasonal temperature cycle on phosphate diffusion and sediment forcing are described by a trigonometric function (Arhonditsis *et al.*, 2007). Phosphorus release from the sediments in the three spatial segments was represented by normal probability distributions, founded upon estimates from previous studies in the Harbour (Mayer and Manning, 1990; Azcue *et al.*, 1998), which were then independently updated by the Bayesian calibration exercise.

References

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

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

Arhonditsis, G.B., Qian, S.S., Stow, C.A., Lamon, E.C., Reckhow, K.H., 2007. Eutrophication risk assessment using Bayesian calibration of process-based models: Application to a mesotrophic lake. *Ecol. Model*. **208**, 215-229.

Azcue, J.M., Zeman, A.J., Mudroch, A., Rosa, F. Patterson, T., 1998. Assessment of sediment Harbour, Canada. *Water Sci. Techol*. **37**, 323-329.

Dermott, R., Johannsson, O., Munawar, M., Bonnell, R., Bowen, K., Burley, M., Fitzpatrick, M., Gerlofsma, J., Niblock, H., 2007. Assessment of lower food web in Hamilton Harbour, Lake Ontario, 2002 - 2004. *Can. Tech. Rep. Fish. Aquat. Sci*. **2729**: 120 p.

Eppley, R.W., Peterson, B.J., 1979. Particulate organic-matter flux and planktonic new production in the deep ocean. *Nature* **282**, 677-680.

Fasham, M.J.R., Ducklow, H.W., McKelvie, S.M., 1990. A nitrogen based model of plankton dynamics in the oceanic mixed layer. *J. Mar. Res*. **48**, 591-639.

Gudimov, A., Stremilov, S., Ramin, M., Arhonditsis, G.B., 2010. Eutrophication risk assessment in Hamilton Harbour: System analysis and evaluation of nutrient loading scenarios. *J. Great Lakes Res.* **36**, 520-539.

Gudimov, A., Ramin, M., Labencki, T., Wellen, C., Shelar, M., Shimoda, Y., Boyd, D., Arhonditsis, G.B., 2011. Predicting the response of Hamilton Harbour to the nutrient loading reductions, A modeling analysis of the "ecological unknowns". *J. Great Lakes Res.* **37**, 494-506.

Hiriart-Baer, V.P., Milne, J., Charlton, M.N., 2009. Water quality trends in Hamilton Harbour: two decades of change in nutrients and chlorophyll a. *J. Great Lakes Res.* **35**, 293-301.

Law, T., Zhang, W., Zhao, J., Arhonditsis, G.B. 2009. Structural changes in lake functioning induced from nutrient loading and climate variability. *Ecol. Model.* **220**, 979-997.

Mayer, T., Manning, P.G., 1990. Inorganic contaminants in suspended solids from Hamilton Harbour. *J. Great Lakes Res.* **16**, 299-318.

Ramin, M., Stremilov, S., Labencki, T., Gudimov, A., Boyd, D., Arhonditsis, G.B., 2011. Integration of mathematical modeling and Bayesian inference for setting water quality criteria in Hamilton Harbour, Ontario, Canada. *Environ Modell Softw.* **26**, 337-353.

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

Zhao, J., Ramin, M., Cheng, V., Arhonditsis, G.B., 2008. Competition patterns among phytoplankton functional groups: How useful are the complex mathematical models? *Acta Oecol* **22**, 324-344.

2. Dataset description

A comprehensive water quality monitoring program with regular sampling of the central area of the Hamilton Harbour has been in place since 1987. For the purposes of our calibration exercise, data from 1987 to 2007, inclusive, with the exception of 1993 during which time no samples were collected, were used to derive the average seasonal patterns in the Harbour. Samples from 1 m depth and the hypolimnion (19 to 22 m depth) were consistently collected during this time period. Nutrient analyses involved measurements of total phosphorus (*TP*), soluble reactive phosphorus (*SRP*), total ammonia (NH_3), nitrate/nitrite ($NO_{3/2}$) and chlorophyll (*chl a*). All these analyses were carried out by the National Laboratory for Environmental Testing in Burlington, Ontario (Environment Canada, 1994). Phosphorus concentrations were determined by colorimetry (chloride–molybdate complex) on unfiltered and filtered lake water samples following acidic persulfate digestion. Total ammonia concentrations were determined by colorimetry (indophenol blue) on filtered lake water samples and $NO_{3/2}$ concentrations were determined by colorimetry (azo dye) following a copper–cadmium column reduction (American Public Health Association, 2005). Chlorophyll concentrations were determined by spectrophotometry following an acetone extraction (Unesco, 1969). Detailed information about the analytical procedures followed and the temporal/spatial trends in the Harbour can be found in the Hiriart-Baer et al. (2009) study. Here, we present a summary of the seasonal variability of the water quality variables considered by the two mathematical models (Table SI-1).

Additional data for our analysis include the solar radiation, day length, precipitation, evaporation based on meteorological data from Environment Canada; namely, the Canadian Daily Climate Data (1996-2002) and the Canadian Climate Normals (1971-2000) (http://www.climate.weatheroffice.ec.gc.ca/ prods_servs/index_e.html). Loads of inorganic nutrients and organic matter enter the Hamilton Harbour from the following main sources: Red Hill and Grindstone creeks, combined sewer overflows (CSOs), Dofasco and Stelco steel mills, Woodward and Skyway wastewater treatment plants (WWTPs), and Cootes Paradise. Estimates of flow and nutrient loadings are based on available data from the Water Survey of Canada (http://www.wsc.ec.gc.ca/) and the RAP loading report

(Hamilton Harbour Technical Team: 1996-2002 Contaminant Loadings and Concentrations to Hamilton Harbour or HHTT-CLR, 2004). The exchanges between the Hamilton Harbour and the relatively high quality waters of Lake Ontario through the Burlington Ship Canal are another major regulatory factor of the Harbour water quality associated with the dilution of the pollutant concentrations, the reduction of Harbour's residence time, and the oxygenation of the hypolimnetic waters (Barica, 1989; Hablin and He, 2003). In particular, the winter exchanges are primarily driven by short-term oscillations due to water level differences at the two ends of the canal, while the exchanges during the summer stratified period are mediated by slowly fluctuating density gradients, i.e., warm Harbour water flowing to the lake in the top layer and colder lake water flowing to the Harbour in the bottom layer (see Figs 1-2 in Barica 1989). Moreover, existing evidence also suggests that the Hamilton Harbour-Lake Ontario interplay during the stratified conditions is much stronger and steadier than the winter period (Hablin and He, 2003). In this study, following the Klabwijk and Snodgrass (1985; see their Fig. 3) conceptual model, we assumed that 20% of the Lake Ontario inflows are directly discharged to the epi- and mesolimnion, whereas 80% of the fresher oxygenated lake water replaces the hypolimnetic masses in the Harbour.

References

American Public Health Association. 2005. Standard methods for the examination of water and wastewater. American Public Health Association, Washington.

Barica, J. 1989. Unique Limnological phenomena affecting water quality of Hamilton Harbour, Lake Ontario. *J. Great Lakes Res*. 15, 519-530.

Environment Canada. 1994. Manual of analytical methods, Volume 1. Major ions and nutrients. Environmental Conservation Service - ECD. Canadian Communications Group., Toronto, ON.

Gudimov, A., Stremilov, S., Ramin, M., Arhonditsis, G.B., 2010. Eutrophication risk assessment in Hamilton Harbour: System analysis and evaluation of nutrient loading scenarios. *J. Great Lakes Re*s. **36**, 520-539.

Hamblin, P.F., He, C., 2003. Numerical models of the exchange flows between Hamilton Harbour and Lake Ontario. *Can. J. Civ. Eng*. **30**, 168-180.

Hamilton Harbour RAP Technical Team (HH RAP). 2004. 1996-2002 Contaminant Loadings and Concentrations to Hamilton Harbour.

Hiriart-Baer, V.P., Milne, J., Charlton, M.N., 2009. Water quality trends in Hamilton Harbour: two decades of change in nutrients and chlorophyll a. *J. Great Lakes Res.* **35**, 293-301.

Klapwijk, A., Snodgrass, W.J. 1985. Model for Lake-Bay Exchange Flow. *J. Great Lakes Res*. **11**, 43-52.

Unesco. 1969. Determination of photosynthetic pigments in sea-water. Imprimerie Rolland-Paris, Paris, France.

Table SI-1: Long term temporal trends of selected water quality variables at the central area of the Hamilton Harbour during the 1987-2007 period.

Depth	*Parameter*	*Winter*	*Spring*	*Summer*	*Fall*
1 m	TP (μg L^{-1})	31.1 ± 21.6	31.6 ± 20.1	36.8 ± 12.6	30.8 ± 9.1
	SRP (μg L^{-1})	11.3 ± 15.8	1.6 ± 8.2	1.5 ± 5.3	2.1 ± 4.9
	$NO_{3/2}$ (mg L^{-1})	2.0 ± 0.2	1.9 ± 0.4	2.0 ± 0.4	1.6 ± 0.3
	$NH_{3\text{-}Tot}$ (mg L^{-1})	0.4 ± 1.3	0.9 ± 0.4	0.2 ± 0.4	0.2 ± 2.8
	Chl *a* (μg L^{-1})	5.5 ± 3.7	7.6 ± 6.6	14.1 ± 8.6	8.9 ± 14.4
	SD (m)	3.1 ± 0.9	2.3 ± 0.6	2.1 ± 0.7	2.5 ± 0.6
> 18 m	TP (μg L^{-1})	30.4 ± 27.1	27.8 ± 18.6	25.4 ± 11.0	31.5 ± 25.6
	SRP (μg L^{-1})	11.4 ± 15.6	1.6 ± 9.3	4.2 ± 5.7	5.1 ± 15.0
	$NO_{3/2}$ (mg L^{-1})	2.0 ± 0.2	1.9 ± 0.4	1.5 ± 0.5	1.1 ± 0.5
	NH3-Tot (mg L^{-1})	0.5 ± 1.3	0.9 ± 0.4	0.2 ± 0.4	0.3 ± 1.6

TP: total phosphorus; SRP: soluble reactive phosphorus; $NO_{3/2}$: nitrate/nitrite; $NH_{3\text{-}Tot}$: total ammonia; chl *a*: chlorophyll *a*; SD: Secchi disc depth.

Table SI-2: State variables of the two eutrophication models.

Model	*Symbol*	*Definition*	*Units*
Simple	*PO4*	Phosphate concentration	µg P L^{-1}
	PHYT	Phytoplankton biomass	µg C L^{-1}
	ZOOP	Zooplankton biomass	µg C L^{-1}
	DET	Detritus concentration	µg P L^{-1}
Complex	*PO4*	Phosphate concentration	µg P L^{-1}
	OP	Organic phosphorus concentration	µg P L^{-1}
	NO3	Nitrate concentration	mg N L^{-1}
	NH4	Ammonium concentration	mg N L^{-1}
	ON	Organic nitrogen concentration	mg N L^{-1}
	CYA	Cyanobacteria biomass	µg C L^{-1}
	PHYT	Phytoplankton biomass	µg C L^{-1}
	ZOOP	Zooplankton biomass	µg C L^{-1}
	DET	Detritus concentration	µg C L^{-1}
	P_{CYA}	Cyanobacteria intracellular phosphorus	µg P µg C^{-1}
	P_{PHYT}	Phytoplankton intracellular phosphorus	µg P µg C^{-1}
	$PO4_{SED}$	Phosphate concentration in the sediments	mg P m^{-2}
	$NH4_{SED}$	Ammonium concentration in the sediments	mg N m^{-2}
	$NO3_{SED}$	Nitrate concentration in the sediments	mg N m^{-2}

Table SI-3: Description and values of the parameters that were not considered during the Bayesian calibration of the complex eutrophication model.

Symbol	*Description*	*Values*	*Units*
$\alpha_{DOC\ zoop}$	Fraction of zooplankton mortality becoming dissolved organic carbon	0.5	-
$\alpha_{DOC\ PHYT}$	Fraction of phytoplankton mortality becoming dissolved organic carbon	0.5	-
$\alpha_{DOC\ CY}$	Fraction of cyanobacteria mortality becoming dissolved organic carbon	0.5	-
$\alpha NH_{4\ zoop}$	Fraction of zooplankton mortality becoming ammonium	0.5	-
$\alpha NH_{4\ PHYT}$	Fraction of phytoplankton mortality becoming ammonium	0.5	-
$\alpha NH_{4\ CY}$	Fraction of cyanobacteria mortality becoming ammonium	0.5	-
αs_{NO3}	Sediment nitrate release rate	0.8	day^{-1}
αs_{PO4}	Sediment phosphate release rate	0.8	day^{-1}
αs_{NH4}	Sediment ammonium release rate	0.8	day^{-1}
$asfood_{det}$	Zooplankton assimilation efficiency for detritus	0.45	-
$asfood_{PHYT}$	Zooplankton assimilation efficiency for phytoplankton	0.5	-
$asfood_{CY}$	Zooplankton assimilation efficiency for cyanobacteria	0.15	-
$ChlaC_{PHYT}$	Chlorophyll to carbon ratio in phytoplankton	0.02	µg Chla µg C^{-1}
$ChlaC_{CY}$	Chlorophyll to carbon ratio in cyanobacteria	0.02	µg Chla µg C^{-1}
$Denitrifmax_{sed}$	Maximum denitrification rate in the sediments	25	mg N m^{-2} day^{-1}
KHdodenit	Half saturation concentration of DO deficit required for nitrification	0.5	mg O_2 m^{-3}
$KHdodenit_{sed}$	Half saturation concentration of DO deficit required for denitrification in the sediments	1	mg O_2 m^{-3}
KHdonit	Half saturation concentration of DO required for nitrification	1	mg O_2 m^{-3}
$KHdonit_{sed}$	Half saturation concentration of DO required for nitrification in the sediments	2	mg O_2 m^{-3}
KHnh4nit	Half saturation concentration of ammonium required for nitrification	1	mg N m^{-3}
$KHnh4nit_{sed}$	Half saturation concentration of ammonium required for nitrification in the sediments	75	mg N m^{-3}
KHno3denit	Half saturation concentration of nitrate required for denitrification	15	mg N m^{-3}
$KHno3denit_{sed}$	Half saturation concentration of DO deficit required for denitrification in the sediments	15	mg O_2 m^{-3}
kt	Effects of temperature on plankton mortality	0.069	$^oC^{-1}$
ktfilt	Effects of temperature on phytoplankton filtration	0.069	$^oC^{-1}$

Symbol	*Description*	*Values*	*Units*
$KTFmin$	Effects of temperature on mineralization	0.004	$^oC^{-2}$
$KTgrdenitr$	Effect of temperature on denitrification	0.004	$^oC^{-2}$
$KTgrdenitr_{sed}$	Effect of temperature on sediment denitrification	0.004	$^oC^{-2}$
$KTgr_{zoop}$	Effect of temperature on zooplankton	0.006	$^oC^{-2}$
$KTgrnitr$	Effect of temperature on nitrification	0.004	$^oC^{-2}$
$KTgrnitr_{sed}$	Effect of temperature on sediment nitrification	0.004	$^oC^{-2}$
$KTgr_{PHYT}$	Effect of temperature on phytoplankton	0.005	$^oC^{-2}$
$KTgr_{CY}$	Effect of temperature on cyanobacteria	0.005	$^oC^{-2}$
kt_{sed}	Effects of temperature on sedimentation	0.004	$^oC^{-1}$
N/C_{zoop}	Nitrogen to carbon ratio for zooplankton	0.2	mg N mg C^{-1}
$Nitrifmax_{sed}$	Maximum nitrification rate in the sediments	50	mg N m^{-2} day^{-1}
P/C_{zoop}	Phosphorus to carbon ratio for zooplankton	0.025	mg P mg C^{-1}
$Pmax_{PHYT}$	Maximum phytoplankton internal phosphorus	0.025	mg P mg C^{-1}
$Pmax_{CY}$	Maximum cyanobacteria internal phosphorus	0.025	mg P mg C^{-1}
$Pmin_{PHYT}$	Minimum phytoplankton internal phosphorus	0.008	mg P mg C^{-1}
$Pmin_{CY}$	Minimum cyanobacteria internal phosphorus	0.008	mg P mg C^{-1}
$Pref_{det}$	Preference of zooplankton for detritus	0.4	-
$Pref_{PHYT}$	Preference of zooplankton for phytoplankton	0.4	-
$Pref_{CY}$	Preference of zooplankton for cyanobacteria	0.2	-
$Tempref$	Reference temperature in the water column	20	^oC
$Tempref_{sed}$	Reference temperature in the sediments	20	^oC
$Toptdenitr$	Optimal temperature for denitrification	20	^oC
$Toptdenitr_{sed}$	Optimal temperature for denitrification in the sediments	20	^oC
$Topt$	Reference temperature for zooplankton	20	^oC
$Toptmin$	Optimal temperature for mineralization	20	^oC
$Toptnitr$	Optimal temperature for nitrification	20	^oC
$Toptnitr_{sed}$	Optimal temperature for denitrification in the sediments	20	^oC
$Topt_{PHYT}$	Reference temperature for phytoplankton metabolism	20	^oC
$Topt_{CY}$	Reference temperature for cyanobacteria metabolism	24	^oC
ψ	Strength of the ammonium inhibition for nitrate uptake	0.05	$(\mu g\ N/L)^{-1}$

FIGURES LEGENDS

Figure SI-1: Comparison between the observed data (black dots) and the median predictions of the *NPZD* model for total phosphorus, phosphate, chlorophyll *a* concentrations, and total zooplankton biomass in the Hamilton Harbour. The 95% credible intervals (dashed lines) stem from the uncertainty pertaining to the model parameters.

Figure SI-2: Comparison between the observed data (black dots) and the median predictions of the complex eutrophication model for total phosphorus, phosphate, ammonium, nitrate, chlorophyll *a* concentrations, total zooplankton biomass, and the abundance of the two phytoplankton functional groups in the Hamilton Harbour. The 95% credible intervals represent the uncertainty pertaining to the model parameters. The grey colour lines and dots in the bottom panel represent the model predictions and the observed biomass of the cyanobacteria-like group.

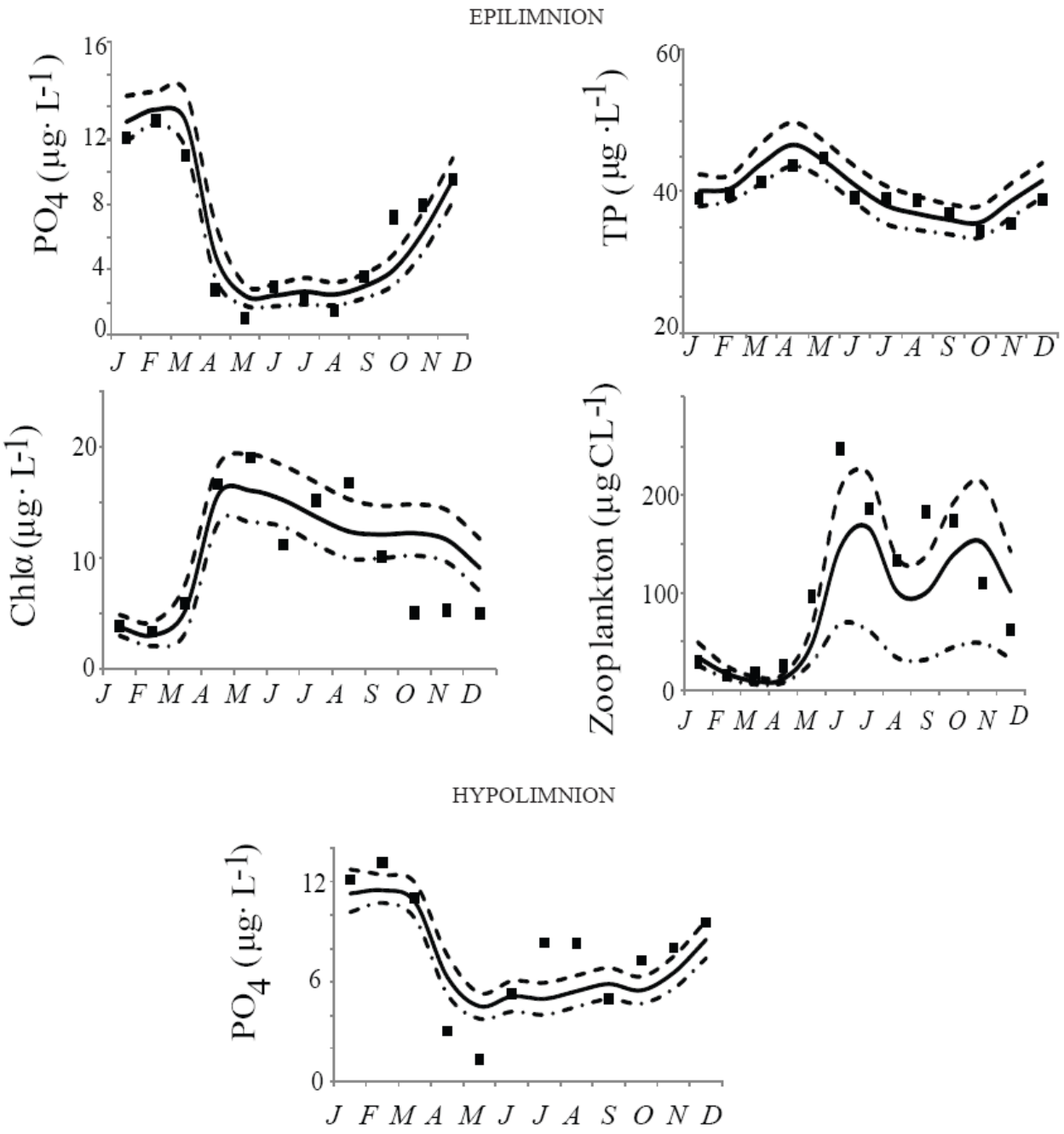
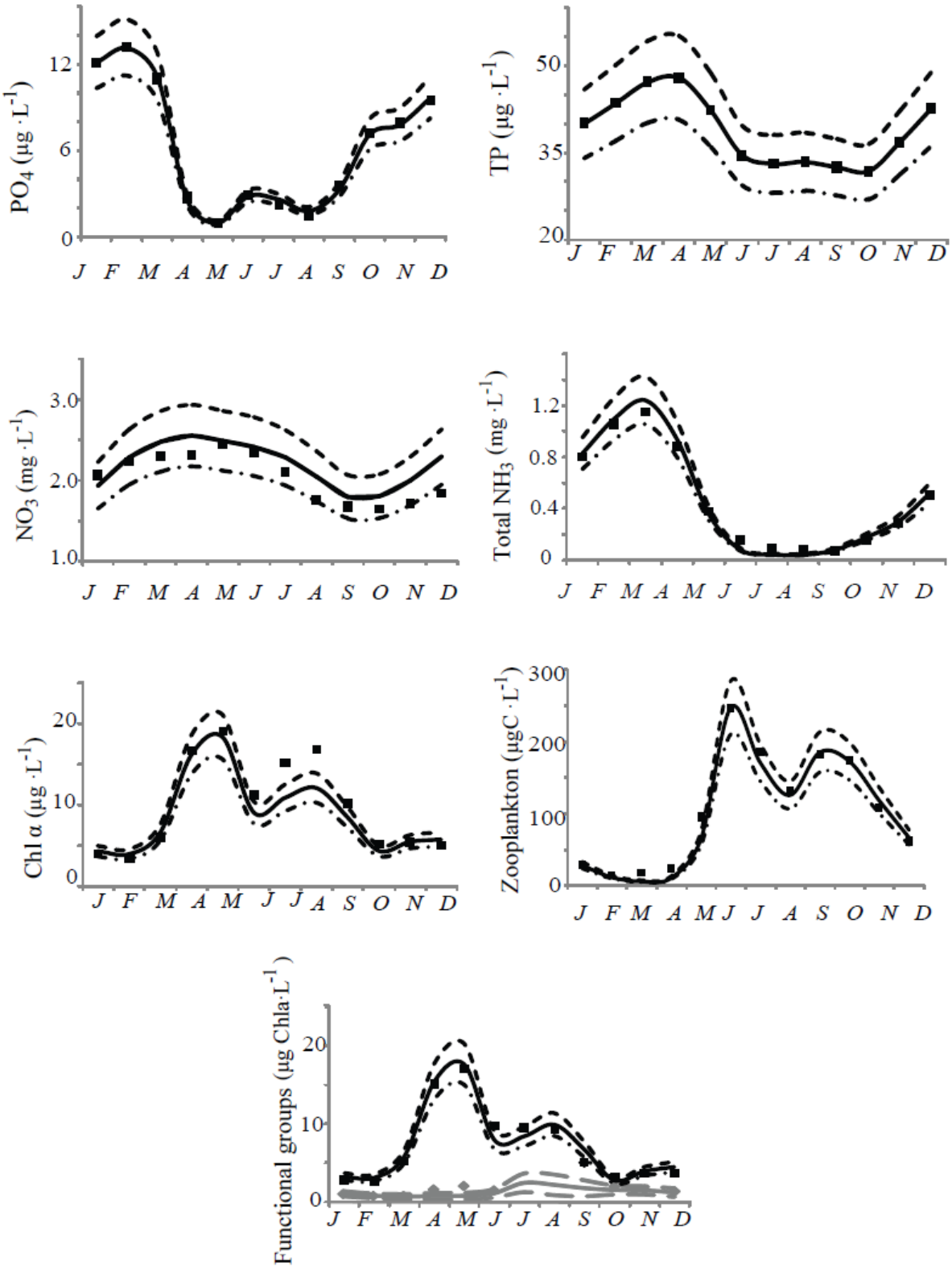

EPILIMNION
PO4 (µg· L-1)
16
12
8
4
0
J F M A M J J A S O N D
TP (µg ·L-1)
60
40
20
J F M A M J J A S O N D
Chlα (µg· L-1)
20
10
0
J F M A M J J A S O N D
Zooplankton (µgCL-1)
200
100
0
J F M A M J J A S O N D
HYPOLIMNION
PO4 (µg· L-1)
12
6
0
J F M A M J J A S O N D

Figure SI-1

EPILIMNION


PO4 (µg ·L-1)
12
6
0
J F M A M J J A S O N D
TP (µg ·L-1)
50
35
20
J F M A M J J A S O N D
NO3 (mg ·L-1)
3.0
2.0
1.0
J F M A M J J A S O N D
Total NH3 (mg ·L-1)
1.2
0.8
0.4
0
J F M A M J J A S O N D
Chl α (µg ·L-1)
20
10
0
J F M A M J J A S O N D
Zooplankton (µgC · L-1)
300
200
100
0
J F M A M J J A S O N D
Functional groups (µg Chla·L-1)
20
10
0
J F M A M J J A S O N D

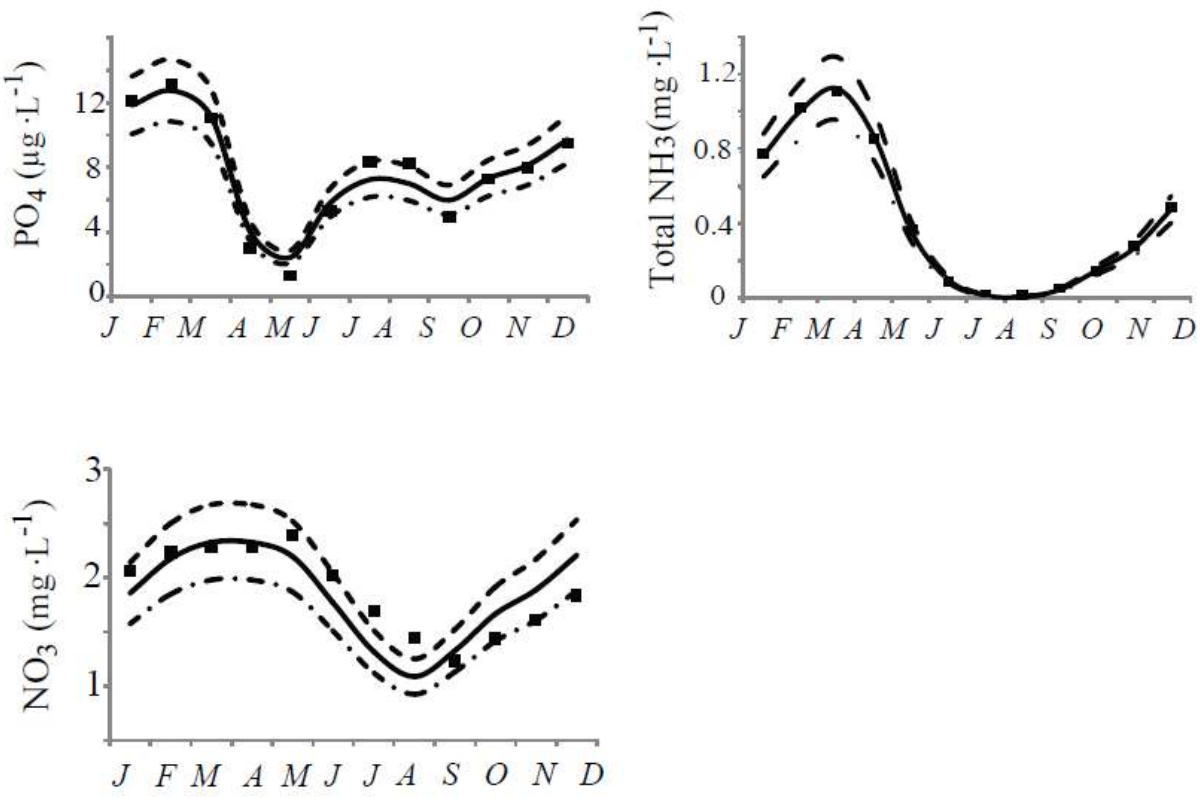

HYPOLIMNION
PO_4 ($\mu g \cdot L^{-1}$)
12
8
4
0
J F M A M J J A S O N D
Total NH_3 ($mg \cdot L^{-1}$)
1.2
0.8
0.4
0
J F M A M J J A S O N D
NO_3 ($mg \cdot L^{-1}$)
3
2
1
J F M A M J J A S O N D

Figure SI-2