Journal of Great Lakes Research 40 Supplement 3 (2014) 73–87

Contents lists available at ScienceDirect

Journal of Great Lakes Research

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

Towards the development of integrated modelling systems in aquatic biogeochemistry: a Bayesian approach

Dong-Kyun Kim [a], Weitao Zhang [a], Veronique Hiriart-Baer [b], Christopher Wellen [a], Tanya Long [c], Duncan Boyd [c], George B. Arhonditsis [a,*]

[a] Ecological Modelling Laboratory, Department of Physical & Environmental Sciences, University of Toronto, Toronto, Ontario, M1C 1A4, Canada
[b] Aquatic Ecosystem Management Research Division, Water Science and Technology Directorate, Science and Technology Branch, National Water Research Institute, Environment Canada, Burlington, Ontario, L7R 4A6, Canada
[c] Great Lakes Unit, Water Monitoring & Reporting Section, Ontario Ministry of the Environment, Environmental Monitoring and Reporting Branch, Toronto, Ontario, M9P 3V6, Canada

ARTICLE INFO

Article history:
Received 26 November 2013
Accepted 21 February 2014
Available online 10 May 2014

Communicated by Craig Stow

Index words:
Process-based modelling
Eutrophication
Bayesian inference
Retrospective analysis
Water quality criteria
Decision making

ABSTRACT

Modelling constructs are designed to shed light on different facets of biogeochemical cycles, but their application involves substantial uncertainty contributed by model structure, parameters, and other inputs. The Bayesian paradigm is uniquely suitable for developing integrated environmental modelling systems, overcoming the conceptual or scale misalignment between processes of interest and supporting information, and exploiting disparate sources of information that differ with regards to the measurement error and resolution. A network of models is developed to connect the watershed processes with the dynamics of the receiving waterbody in the Hamilton Harbour (Ontario, Canada). The SPAtially Referenced Regressions On Watershed attributes (*SPARROW*) along with an intermediate complexity eutrophication model were used to reproduce the phosphorus cycling in the system, including the exchange between sediment and water column as well as the interplay between the ambient and phytoplankton intracellular pools. The novel features of the framework include (*i*) the development of a downscaling algorithm that transforms the *SPARROW* annual phosphorus loading estimates to daily inputs for the eutrophication model; and (*ii*) a neural network that emulates the posterior linkages between model parameters/phosphorus loading inputs and the predicted total phosphorus, chlorophyll *a* concentrations, and zooplankton abundance. Our integrated watershed-receiving waterbody model is independently tested against a 22-year period (1988–2009) and is subsequently used to gain insights into the ecological factors that shape the current water quality conditions in the system and may modulate its future response to the nutrient loading reductions proposed by the Hamilton Harbour Remedial Action Plan.

Introduction

In his 2006 essay, titled "*A manifesto for the equifinality thesis*", K.J. Beven discussed the reasons why many earth science modellers are reluctant to embrace the idea of equifinality and usually strive for the identification of a global optimum in the parameter space that will maximize the model fit to the observed data. Reflecting the popular stance that any scientific endeavour should aspire to achieve a single correct description of the reality, issues related to model parametric uncertainty or even to the adequacy of a model structure are typically downplayed. Any ambiguity in the predictive statements supported by a model is perceived as an undermining of its credibility in guiding management decisions about future investments to the environment (Beven, 2006). Nonetheless, Pappenberger and Beven (2006) challenged several of these skeptical views, arguing that uncertainty analysis is necessary because there are no such models as "physically realistic". Uncertainty analysis can be used for hypothesis testing; uncertainty bounds can be understood by stakeholders and policy makers; uncertainty analysis is not necessarily subjective or difficult to perform; and can be an integral component of the decision-making process. There are indeed several compelling reasons to rigorously quantifying the uncertainty associated with any modelling exercise and effectively communicating the robustness of model projections in policy analysis frameworks (Arhonditsis et al., 2007; Beven and Alcock, 2012). In this regard, one of the emerging imperatives towards the broader adoption of uncertainty analysis in contemporary modelling practice may be the establishment of formal guidelines that will optimize the associated procedures, the so-called "Code of Practice" (sensu Pappenberger and Beven, 2006).

The quantification of uncertainty associated with the multidimensional parameter spaces of mathematical models involves three critical decisions: i) the characterization of the uncertainty underlying the model parameters prior to model calibration; ii) selection of the sampling scheme for generating input vectors which are evaluated in regards to the model performance; and iii) selection of the likelihood measure to quantify model fit to the observed data (Arhonditsis et al.,

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

http://dx.doi.org/10.1016/j.jglr.2014.04.005
0380-1330/

2011). Regarding the former decision, Arhonditsis et al. (2007) argued that there is usually sufficient empirical knowledge (e.g., field observations, laboratory studies, literature information, and expert judgment) to delineate plausible regions of the parameter space of aquatic ecosystem models, and thus effectively reducing the mismatch between what ideally we would like to learn (model structure) and what realistically can be observed (calibration datasets). The selection of the strategy for generating parameter vectors determines the sampling efficiency of the modelling exercise, e.g., Random sampling, Latin hypercube, Markov chain Monte Carlo (*MCMC*). Sampling algorithms that draw samples uniformly and independently from the prior parameter space often insufficiently cover regions of high model likelihood; especially, when the joint prior parameter distribution is very wide or the parameters are highly correlated (Qian et al., 2003). To address this problem, there are propositions to efficiently estimate the posterior probability density function of parameters in complex high dimensional modelling problems, using adaptive *MCMC* schemes that ensure ergodicity while adjusting the scale and orientation of the proposal distributions, e.g., the differential evolution adaptive Metropolis (Vrugt et al., 2008). The selection of the likelihood measures determines whether the model assessment should be based on a generalized (e.g., Root Mean Square Error, U-uncertainty, Reliability Index, Modelling Efficiency) or a formal probabilistic (e.g., Normal, Lognormal or Poisson error) likelihood function. The latter decision can significantly alter the inference drawn, and the implications of each choice have been extensively debated in the literature (Beven and Young, 2003; Thiemann et al., 2001).

Viewing model uncertainty analysis as an attempt to formulate the joint probability distribution of model inputs and then update our knowledge about this distribution after the consideration of the calibration dataset, the Bayesian inference represents a suitable means to combine existing information (prior) with current observations (likelihood) for projecting the future. There are several recent studies that attempted to illustrate how Bayesian inference techniques can be used to quantify the information that data contain about model inputs, to offer insights into the covariance structure among parameter estimates, and to obtain predictions along with uncertainty bounds for model outputs (Bayarri et al., 2007; Higdon et al., 2004; Kennedy and O'Hagan, 2001). Specifically, in the context of water quality modelling, Bayesian calibration schemes have been introduced using simple mathematical models (<10 state variables) and statistical formulations that explicitly accommodate measurement error, parameter uncertainty, and model structure error (Arhonditsis et al., 2007, 2008a,b, 2011). Striving for an improvement of the credibility of model-based water quality management, recent trends have also focused on the capacity of more complex mathematical constructs to be combined with Bayesian calibration techniques (Ramin et al., 2011; Reichert and Schuwirth, 2012; Zhang and Arhonditsis, 2008). In this regard, significant improvements have been made with respect to the associated computational demands and the error propagation control (Dietzel and Reichert, 2012), but the emergence of the holistic management paradigm has increased the demand for even more complex biogeochemical models with considerably greater uncertainty (Zhang and Arhonditsis, 2008). In particular, there is increasing pressure for the development of integrated water quality models that effectively connect watershed with downstream biogeochemical processes. This need stems from the emerging management questions related to contemporary climate and land use changes (Rode et al., 2010). Modelling nutrient transport and water quality dynamics is challenging due a number of constraints associated with input data as well as existing knowledge gaps related to the mathematical depiction of the land-to-water delivery, in-stream attenuation and eutrophication processes in the receiving waterbody.

To this end, our goal is to offer an illustrative case study of how the Bayesian approach can assist in developing integrated environmental modelling systems, overcoming the conceptual or scale misalignment between processes of interest and supporting information, and exploiting disparate sources of data that differ with regards to their quality and resolution. In particular, we develop a network of models that aims to connect the watershed processes with the dynamics of the receiving waterbody in the Hamilton Harbour, Ontario, Canada, using the *SPARROW* non-linear regression strategy along with an intermediate complexity process-based model that aims to reproduce the phosphorus cycling in the system (Fig. 1). The novel features of our framework include (*i*) the development of a downscaling algorithm that transforms the *SPARROW* annual phosphorus loading estimates to daily inputs for the eutrophication model; and (*ii*) a neural network that emulates the posterior linkages between model parameters/phosphorus loading inputs and the predicted total phosphorus, chlorophyll *a* concentrations, and zooplankton abundance. Our study undertakes a predictive confirmation of the integrated modelling framework against a 22-year period (1988–2009) and subsequently attempts to shed light on critical aspects of the system dynamics that invite further investigation and will likely determine our predictive capacity to assess compliance with the existing delisting water quality targets.

Methodology

Study site

Our study site, Hamilton Harbour, is a 21.5 km^2 embayment located in the western end of Lake Ontario with drainage basin of about 450 km^2 (Fig. 2). The hydraulic load of the Harbour is mainly contributed by municipal and industrial sewage effluents (e.g., wastewater treatment plants, combined sewer overflows, and steel mills) and stream discharges which collect urban (e.g., Redhill Creek) and agricultural (e.g., Grindstone Creek) runoff (Hiriart-Baer et al., 2009). The Harbour has restricted water exchange with Lake Ontario via the Burlington Ship Canal (Rao et al., 2009). Hamilton Harbour experiences serious water quality issues, such as blooms of undesirable algae, low water clarity, hypoxia during the summer, bacterial contamination, as a consequence of excessive loadings of nutrients and other pollutants. Since the mid 1980s, when the Harbour was identified as one of the 43 Areas of Concern (*AOC*) in the Great Lakes area, the Hamilton Harbour Remedial Action Plan (*RAP*) was formulated through a variety of government, private sector, and community participants to provide the framework for actions aimed at restoring the Harbour environment. The foundation of the remedial measures and the setting of water quality goals reflect an ecosystem-type approach that considers the complex interplay between abiotic variables and biotic components pertinent to its beneficial uses (Charlton, 2001). Drastic nutrient loading reduction has historically played a central role in the restoration efforts, although the determination of the critical levels has been a contentious issue as the population growth and increasing urbanization accentuate the pressure for expansion of the local wastewater treatment plants (WWTPs). Recent modelling work also suggests that the water quality goals for *TP* levels <20 μg L^{-1}, chlorophyll *a* concentrations between 5–10 μg L^{-1}, and water clarity >3 m will likely be met, if the proposed phosphorus loading reductions to 142 kg day^{-1} are actually achieved and there is an allowable 10% frequency of violations (Gudimov et al., 2010, 2011; Ramin et al., 2011, 2012). Nonetheless, it was emphasized that the predictive capacity of any modelling exercise in the system is conditional upon the credibility of the contemporary nutrient loading estimates, which are uncertain and appear to inadequately account for the contribution of non-point sources, episodic meteorological events (e.g., spring thaw, intense summer storms), and short-term variability at the local *WWTP*s (Wellen et al., 2012, 2014).

Integrated modelling framework

We developed an integrated modelling network that is founded upon (*i*) a *SPARROW* model configuration that accommodates the inter-annual phosphorus loading variability in the Hamilton Harbour

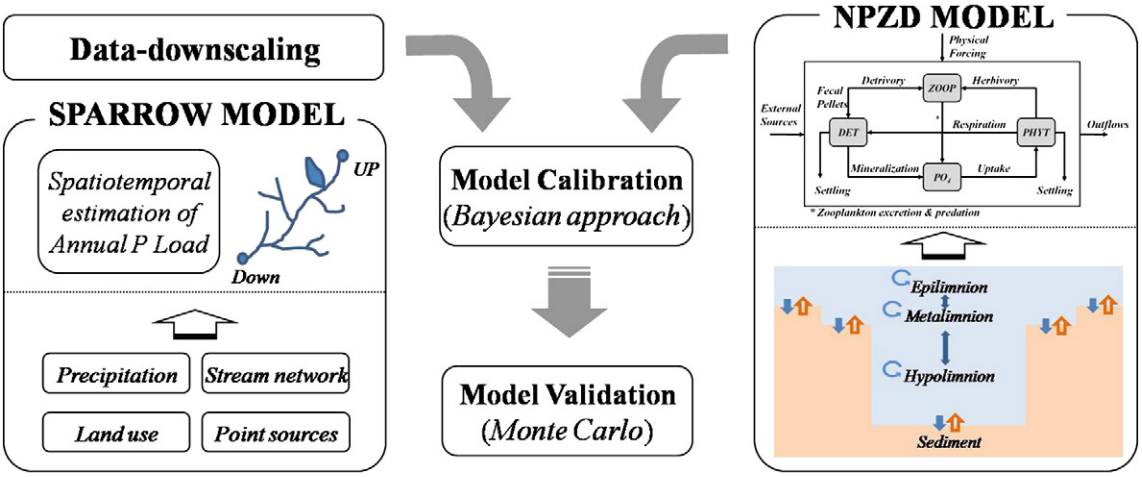

Fig. 1. Graphical representation of the integrated biogeochemical model developed for the Hamilton Harbour. NPZD denotes the model for the receiving waterbody that considers the interplay among the limiting nutrient (phosphate), phytoplankton, zooplankton, and detritus (particulate phosphorus).

watershed; (*ii*) a Bayesian downscaling algorithm that transforms the annual phosphorus loading predictions to daily estimates; and (*iii*) a eutrophication model that aims to reproduce the interplay among phosphorus cycling, phytoplankton dynamics, and herbivorous grazing (Fig. 1).

SPARROW model

SPARROW model was developed by the United States Geological Survey (*USGS*) to estimate nutrient loads, yields, and deliveries at landscape and regional scales (McMahon et al., 2003). The model uses a hybrid

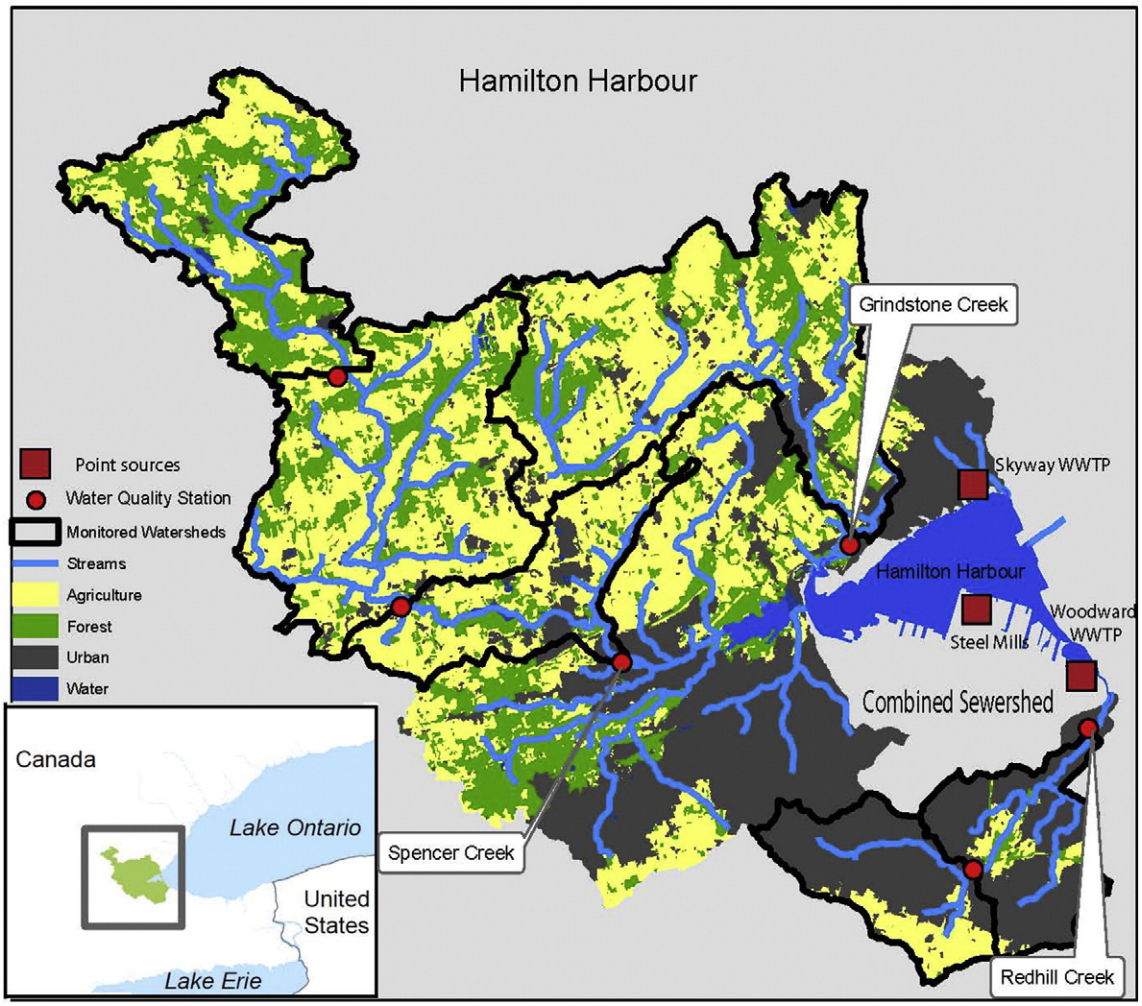

Fig. 2. Map of the Hamilton Harbour watershed, western end of Lake Ontario, Ontario, Canada, its land use, point sources and sampling stations.

empirical/process-based approach designed to be applied to a network of water quality monitoring stations. *SPARROW* consists of a two-level hierarchical spatial structure. Watersheds are first divided into subwatersheds, each of which drains to a water quality monitoring station, and then each subwatershed is disaggregated into reach catchments draining to a particular stream segment (Schwarz et al., 2006). *SPARROW* considers two basic processes: (*i*) the source processes, represented by export coefficients, accounting for the constituent mobilization; (*ii*) the sink processes introduced by delivery factors, predicting how landscape attributes modulate the delivery of the mobilized constituent to streams, and attenuation coefficients, predicting the amount of the delivered constituent remaining in transit per length of stream or per reservoir. Bayesian parameter estimation has been used for addressing several core issues related to the *SPARROW* applications; (*i*) the uncertainty of calibration data, (*ii*) the importance of informative prior parameter distributions in assisting model calibration, (*iii*) the characterization of the spatial structure of model residuals due to autocorrelated forcing factors (e.g., climate and soils) or inadequacies of the model structure (e.g., a single export coefficient for all the agricultural land uses may overestimate the intensity of the agricultural practices in certain sub-watersheds and underestimate them in others), and (*iv*) the implications of the covariance of model parameters on the inference drawn and the posterior patterns derived (Qian et al., 2005; Wellen et al., 2014).

The *SPARROW* approach typically expresses the average annual nutrient loads hydrologically de-trended to a particular base year, and thus the focus is placed on the spatial heterogeneity of the nutrient sources in the studied watershed and the influence of hydrological transport is minimized. To accommodate the year-to-year phosphorus loading variability, Wellen et al. (2012) introduced the *SWALLOW* (SPARROW *W*ith *A*nnua*L* *L*oads *O*f *W*atersheds) framework, comprising two major strategies, in which: (*i*) the *SPARROW* model provides a static baseline, i.e., a long-term annual load estimate, while the inter-annual variability is captured by linear expressions of climatic variables (e.g., precipitation, evapotranspiration); and (*ii*) the source/sink processes within the *SPARROW* model are allowed to vary at annual timescales using dynamic parameter estimation techniques akin to those used in dynamic linear models. In the present study, we employed the latter approach without the consideration of any climatic forcing factors. Specifically, the annual log-transformed nutrient loadings were assumed to be a draw from a normal distribution with a mean defined by the *SWALLOW* model, a constant model (process) error variance, and measurement error terms independent both in space and time:

$$Y_{i,t} \sim N\left(Load_{i,t}, {\delta_{i,t}}^2\right)$$

$$Load_{i,t} \sim N\left(\mu_{i,t}, \sigma^2_{SPARROW}\right)$$

$$\mu_{i,t} = \ln\left(\sum_{n=1}^{Ns}\sum_{j=1}^{J_i}\beta n, tSn, je\left(-\alpha Z_j\right)Hi, j, tSHi, jR\right)$$

$$H^S_{i,j,t} = \prod_m \exp\left(-k_{s,m,t}L_{i,j,m}\right)$$

$$H^R_{i,j} = \prod_l \exp\left(-k_r q_l^{-1}\right)$$

$$\sigma^{-2}_{SPARROW} \sim gamma(0.001, 0.001)$$

where $Y_{i,t}$ refers to the natural logarithm of the measured annual load at subwatershed monitoring station i during year t (tons P yr^{-1}); $Load_{i,t}$ is a latent variable that represents the "true" loading values when accounting for the measurement error $\delta_{i,t}$; $\mu_{i,t}$ is a prediction of the natural logarithm of the annual load at monitoring station i for year t estimated by the *SPARROW* model; n and N_s refer to the source index, where N_s is the total number of sources (diffuse and point sources) and n is an index for each source; J_i refers to the number of reaches in subwatershed i; $\beta_{n,t}$ refers to the estimated source coefficient for source n (tons P km^{-2} yr^{-1} for non-point sources) in year t; $S_{n,j}$ refers to the quantity of source n in reach j; km^2 of agricultural or urban land-use area for non-point sources; metric tons yr^{-1} for point sources; α refers to the global coefficient for land to stream delivery attenuated by wetlands; Z_j is a vector of the wetland areas associated with drainage in reach j; $H^S_{i,j}$ refers to the fraction of nutrient mass originating in reach j remaining at station i as a function of a year-specific first-order loss processes in streams; $H^R_{i,j}$ refers to the fraction of nutrient mass originating in reach j remaining at station i as a function of first order loss processes in lakes and reservoirs; $k_{s,m,t}$ refers to the first order loss coefficient for stream class m (km^{-1}) in year t; $L_{i,j,m}$ refers to the class m stream length in kilometers between reach j and station i; l refers to the number of lakes or reservoirs between reach j and station i; k_r refers to the first order loss coefficient or settling velocity (m year^{-1}); q_l refers to the aerial hydraulic loading (m year^{-1}) of any of the l lakes/reservoirs between reach j and station i; and $\sigma^2_{SPARROW}$ represents variance indicative of the model (process/structure) error, a draw from the gamma distribution with shape and scale parameters of 0.001, representing a "non-informative" (or vague) prior assigned to the error precision (the inverse of variance). The measurement error (${\delta_{i,t}}^2$) terms were pre-specified from the 95% confidence intervals of the daily loads calculated by the rating curve model, as provided by the LOADEST program (Runkel et al., 2004):

$${\delta_{i,t}}^2 = \sum_{x,y} Cov\left(Load_{i,t,x}, Load_{i,t,y}\right)$$

where $Load_{i,t,x}$ and $Load_{i,t,y}$ denote the loads on the arbitrary days x and y, and thus the terms ${\delta_{i,t}}^2$ correspond to the variance of the mean predicted loads, estimated as the sum of the covariance of all the predicted daily loads at station i for year t. The covariance terms were based on the equations given by Gilroy et al. (1990, Equations 17–25), and account for the residual variance of the rating curve model along with the parametric uncertainty.

As previously mentioned, four parameters were treated as time dependent: namely, the agricultural export coefficient (β_1), the urban export coefficient (β_2), and the attenuation coefficients for small (ks_1) and large (ks_2) streams. Our dynamic approach postulates that the time-variant parameters are serially correlated, i.e., adjacent parameter values are more similar relative to those that are distant in time. Counter to the conventional regression analysis, in which parameters are conditioned upon the entire time series, the dynamic parameter estimation is influenced only by prior and current information, not by subsequent data (Sadraddini et al., 2011a,b). Thus, the influence of the original priors decreases as time progresses and is gradually superseded by the influence of the data. In particular, we assumed that 98% of the information is carried forward from year t to $t+1$, and the statistical configuration can be expressed as follows:

$$\theta_t \sim N\left(\theta_{t-1}, \phi_t^2\right)\Big|(\theta_{\min}, \theta_{\max})$$

$$\phi_t^{-2} = 0.98^t \cdot \phi_1^{-2}$$

$$\theta_1 \sim N\left(\theta_0, \phi_1^2\right)\Big|(\theta_{\min}, \theta_{\max})$$

$$\phi_1^{-2} \sim gamma(a, \beta)$$

where θ_t represents one of the parameters from the time-varying vector [β_1,β_2,ks_1,ks_2]; ${\varphi_t}^2$ refers to the corresponding time-dependent variance; 0.98 is the discount factor; θ_0 represents the prior mean value assigned to the time-varying parameter for the beginning of our simulation period; $\theta_{\min}$ and $\theta_{\max}$ are the (literature-based) lower and upper bounds used to eliminate any unrealistic draws for θ_1 as well as for the subsequent θ_t values, and ${\varphi_1}^2$ corresponds to the parameter variance for the first simulated year, which was assigned an informative gamma prior (see description in Wellen et al., 2012, page 6). The prior distributions used for both time dependent and independent parameters can be found in Table 1. The calibration dataset for the parameter estimation of the dynamic *SPARROW* includes TP loading measurements

Table 1
Prior and posterior mean and standard deviation values of the *SPARROW* model parameters.

		α		β_1		β_2		k_{s1}		k_{s2}		k_r		σ	
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Priors		0	3.2									15.221	4.757	1.000	5.623
	1988			0.160	1.251	0.258	3.515	0.318	0.231	0.082	0.108				
Posteriors		10.41	1.228									15.330	3.569	0.034	0.011
	1988			0.255	0.068	0.042	0.028	0.231	0.119	0.120	0.042				
	1989			0.160	0.048	0.038	0.022	0.186	0.099	0.083	0.028				
	1990			0.209	0.059	0.075	0.056	0.148	0.087	0.036	0.019				
	1991			0.171	0.060	0.062	0.050	0.142	0.103	0.036	0.023				
	1992			0.191	0.066	0.068	0.045	0.120	0.087	0.030	0.022				
	1993			0.144	0.047	0.070	0.049	0.098	0.083	0.023	0.019				
	1994			0.104	0.036	0.049	0.035	0.101	0.103	0.020	0.018				
	1995			0.125	0.039	0.086	0.064	0.116	0.099	0.014	0.013				
	1996			0.201	0.055	0.095	0.063	0.104	0.078	0.013	0.012				
	1997			0.137	0.038	0.054	0.033	0.094	0.074	0.014	0.012				
	1998			0.108	0.042	0.075	0.050	0.087	0.077	0.024	0.024				
	1999			0.066	0.034	0.059	0.030	0.089	0.079	0.046	0.042				
	2000			0.142	0.046	0.060	0.036	0.079	0.072	0.036	0.027				
	2001			0.112	0.039	0.059	0.033	0.077	0.073	0.032	0.026				
	2002			0.085	0.033	0.041	0.027	0.082	0.088	0.025	0.021				
	2003			0.118	0.046	0.077	0.058	0.087	0.093	0.022	0.020				
	2004			0.124	0.053	0.072	0.070	0.106	0.121	0.022	0.024				
	2005			0.143	0.059	0.078	0.078	0.124	0.137	0.020	0.022				
	2006			0.210	0.117	0.098	0.104	0.136	0.158	0.022	0.027				
	2007			0.107	0.048	0.081	0.081	0.139	0.157	0.021	0.024				
	2008			0.242	0.106	0.108	0.117	0.133	0.151	0.019	0.021				
	2009			0.196	0.091	0.082	0.078	0.150	0.188	0.020	0.022				

for our 22-year study period from 6 sub-watersheds (Wellen et al., 2012). The predicted annual non-point loading values along with the uncertainty estimates for two downstream tributaries (Grindstone and Redhill Creeks) were used to force the eutrophication model for the receiving water body.

Eutrophication model

Our eutrophication model was an augmented version of the simple structure presented by Ramin et al. (2012) to guide the water quality criteria setting process in the Hamilton Harbour. The mathematical structure of the model and its parameters are given in Electronic Supplementary Material (ESM) Tables S1,S2). This section provides the basic conceptual design of the eutrophication model and the features altered in the present study while detailed description of its structure can be found elsewhere (Arhonditsis et al., 2007; Law et al., 2009; Ramin et al., 2012). Our model considers the interplay among the limiting nutrient (phosphate), phytoplankton, zooplankton, and particulate phosphorus. The spatial segmentation of the model consists of three compartments representing the epilimnion, metalimnion, 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 with explicit consideration of the control exerted by the intra-cellular storage practices (i.e., luxury uptake). 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 (Holling, 1959). 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. Zooplankton mortality/predation losses were accounted for by a sigmoidal closure term (Edwards and Yool, 2000). A fraction of the zooplankton grazing is assimilated and fuels growth, another fraction is excreted as phosphate, while the remaining fraction represents the 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. A simple mechanistic approach was used to relate the fluxes of 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 first step of our analysis involved the calibration of the coupled watershed-receiving water model to reproduce the average water quality patterns in the Hamilton Harbour. The data for this training exercise were based on an independent 3-year (2002–2004) dataset collected by the Department of Fisheries and Oceans (Dermott et al., 2007). The prior information on the calibration parameter vector is summarized in Table 2. The statistical formulation used to guide the Bayesian calibration was founded upon the assumption that the eutrophication model is an imperfect simulator of the system dynamics and the structural error is constant over the annual cycle for each state variable (Arhonditsis et al., 2007). The uncertainty associated with the dataset was also accounted for with a data quality submodel (Wellen et al., 2012). In particular, we assumed that the monthly standard deviations of the modelled water quality variables were 25% of the corresponding mean monthly values; a fraction that comprises both analytical error and the inter-annual variability in the Hamilton Harbour (Hiriart-Baer et al., 2009; Ramin et al., 2011). In this phase, we used the static version of the *SPARROW* model presented by Wellen et al. (2014) to provide the average non-point source loading in Hamilton Harbour watershed, which were then downscaled into daily values (see also ESM Table S3).

The second phase of our analysis involved the predictive confirmation of the model over the 22-year study period (1988–2009), based on the joint posterior parameter patterns of the integrated watershed-receiving waterbody model. The posterior estimates of the mean and standard deviation parameter values along with the covariance structure were used to update the eutrophication model (Arhonditsis et al., 2011; Ramin et al., 2012). Under the assumption of a multivariate

Table 2
Description of the calibration parameter vector of the eutrophication model along with the prior and posterior mean and standard deviation values.

Symbol	Description	Units	Priors		Posteriors	
			Mean	S.D.	Mean	S.D.
a	Maximum phytoplankton growth rate	day^{-1}	1.772	0.382	1.327	0.233
α	Zooplankton assimilation efficiency		0.439	0.052	0.352	0.029
β_{PO4}	Zooplankton excretion fraction to phosphate		0.391	0.060	0.352	0.052
β_P	Fraction of refractory phosphorus buried into deeper sediment		0.725	0.087	0.952	0.014
d	Zooplankton mortality rate	day^{-1}	0.114	0.015	0.137	0.014
K_p	Half-saturation constant for PO_4 uptake	mg P m^{-3}	13.01	4.664	10.19	2.903
γ	Zooplankton predation excretion fraction to phosphorus		0.450	0.098	0.298	0.074
I_s	Half-saturation light intensity	MJ m^{-2} day^{-1}	160.6	28.79	150.8	24.53
K_b	Background light extinction coefficient	m^{-1}	0.214	0.029	0.216	0.029
K_c	Light extinction coefficient due to chlorophyll α	L(μg chlα m)$^{-1}$	0.031	0.013	0.045	0.013
λ	Maximum zooplankton grazing rate	day^{-1}	0.571	0.077	0.525	0.050
μ	Zooplankton grazing half-saturation coefficient	mg P m^{-3}	5.297	3.360	6.132	1.085
ω	Relative zooplankton preference for detritus compared to phytoplankton		0.597	0.155	0.035	0.018
φ	Detritus mineralization rate	day^{-1}	0.311	0.176	0.010	0.003
P_{maxup}	Maximum phosphorus uptake rate for phytoplankton	μg P/L^{-1} day^{-1}	0.023	0.008	0.033	0.008
$pred$	Zooplankton predation half-saturation coefficient		53.18	11.47	44.30	8.668
ψ	Detritus sinking rate	m day^{-1}	0.341	0.252	0.478	0.028
r	Phytoplankton respiration rate	day^{-1}	0.035	0.016	0.027	0.009
s	Phytoplankton sinking velocity	m day^{-1}	0.068	0.050	0.055	0.037

normal distribution for the parameter values, the conditional distributions are given by:

$$\hat{\theta}_{i|j} = \hat{\theta}_i + \left[\theta_j - \hat{\theta}_j\right]\Sigma_j^{-1}\Sigma_{i,j}$$
$$\Sigma_{i|j} = \Sigma_i - \Sigma_{j,i}\Sigma_j^{-1}\Sigma_{i,j}$$
$$j \in \{i+1, \ldots n\}$$

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 $\hat{\theta}_i, \hat{\theta}_j, \theta_j$ correspond to the posterior mean and random values of the parameters i and j, respectively. A comprehensive water quality sampling program has been conducted by Environment Canada (EC) in the central area of Hamilton Harbour during ice-free season since 1987 (Hiriart-Baer et al., 2009). The variables used in this study include the vertical temperature profiles, total phosphorus (TP), phosphate (PO_4), and chlorophyll a (chl$a$). More information about the sampling protocol and the analytical methods used can be found in Hiriart-Baer et al. (2009). The external forcing consisted of the hydrological conditions (e.g., inflow, water exchange with Lake Ontario, precipitation, and evaporation), water temperatures, and exogenous phosphorus loading. The total phosphorus loading included historical point source records from the local wastewater treatment plants, industrial discharges, combined sewage overflows, inflows from Cootes Paradise, as well as the non-point source estimates for Grindstone and Redhill Creeks from the *SWALLOW* model (see Section 2.2.1). The year-specific loading estimates were then downscaled to daily resolution inputs using the algorithm described in the following section.

Downscaling algorithm

As previously noted, *SPARROW* model only provides annual nutrient loads, whereas the eutrophication model requires inputs with daily resolution. To address the mismatch between the two models, we developed a Bayesian hierarchical downscaling algorithm. First, we used principal component analysis to identify different seasonal modes of intra-annual variability with respect to the long-term daily flow time series in Redhill and Grindstone Creek subwatersheds (Arhonditsis et al., 2004a,b; Jassby, 1999). Our analysis revealed three independent hydrological seasons within the annual cycle, representing the spring until early summer (April-June), summer until early fall (July-September), late fall and the entire winter (October-March) period. Piecewise regression was used to connect the daily precipitation with the downstream flows. In doing so, we were able to identify threshold precipitation levels that can induce low and high flow conditions in each hydrologic period and creek. Logistic regression models were then developed to predict the likelihood of low and high flow regimes based on daily precipitation values. The actual specification of the flow regime for a particular day was modelled as a Bernoulli draw, driven by the projections of the logistic regression models. Thus, the first phase of the downscaling algorithm can be mathematically expressed as follows:

$$\text{logit}[p_{FRt}] = \alpha + \beta X_t$$
$$X_t = \ln(precipitation_t + 1)$$
$$FR_t \sim bernoulli(p_{FRt})$$
$$i = FR_t + 1 = \begin{cases} 1\ (low\ flow\ regime) \\ 2\ (high\ flow\ regime) \end{cases}$$

where i is the index of the flow regime, FR denotes the flow regime characterization for day t, p_{FRt} represents the likelihood of the flow regime in day t, α and β are the intercept and slope terms of the logistic regression models that connect the daily precipitation with the likelihood p_{FRt}, X_t is the predictor of the likelihood of the flow regime in day t, which was the 2-day moving average of the daily precipitation (logarithmic scale).

In the next step, the log-transformed daily flows within each hydrological regime were fitted to normal distributions, and the derived moments (mean μ_{Flow} and variance σ^2_{Flow}) were subsequently used to generate daily flows and *TP* concentrations. In particular, we statistically reproduced the daily flows and *TP* concentrations by explicitly considering the covariance between the two variables, assuming that they follow a multivariate normal distribution (*MVN*):

$$\begin{pmatrix} Flow_i \\ TP_i \end{pmatrix} \sim MVN\left(\begin{pmatrix} \mu_{Flow_i} \\ \mu_{TP_i} \end{pmatrix}, \Omega_i\right)$$
$$\Omega_i = \begin{pmatrix} \sigma^2_{Flow_i} & \rho_i\sigma_{Flow_i}\sigma_{TP_i} \\ \rho_i\sigma_{Flow_i}\sigma_{TP_i} & \sigma^2_{TP_i} \end{pmatrix}$$
$$\overline{Load_i} = Flow_i \cdot TP_i$$
$$Load_i \sim N\left(\overline{Load_i}, \sigma^2_{SPARROW_{daily}}\right)$$

where $Flow_i$ and TP_i represent the log-transformed daily flows and *TP* concentrations for flow regime i (1 or 2) during each hydrologic period in each creek; μ_{Flowi} and σ^2_{Flowi} are the mean and variance assigned to each flow regime i after fitting the flow data; μ_{TPi} is the annual average

TP concentration calculated by dividing the *SPARROW* load prediction with the annual average flow in each creek; σ^2_{TPi} represents the associated variance of the measured *TP* concentrations; Ω_i is the covariance matrix between flow and *TP*; $Load_i$ is the daily non-point loading used to force our eutrophication model; $\sigma^2_{SPARROWdaily}$ the *SPARROW* process error expressed in daily load terms; and ρ_i is the correlation coefficient between flow and *TP* for the flow regime *i*. In our dataset, the covariance between flow and *TP* was weaker in low rather than in high flow regimes. Thus, the correlation coefficient (ρ) in low flow regimes ($i = 1$) was simulated as a uniform draw from the 0.1-0.2 range, whereas the high flow regimes ($i = 2$) were sampled from the 0.4-0.6 range.

Emulation of the model posterior using neural networks

There is a growing acceptance that complex models may not be easily subject to rigorous uncertainty analysis, because of their need for tens of thousands, if not millions, of model simulations. Statisticians are increasingly turning to model emulators to enable robust sensitivity and uncertainty analyses. While a *simulator* is a model of the real world, an *emulator* is a statistical approximation of a simulator which can effectively overcome the large computational demands associated with the actual process-based models (Castelletti et al., 2012). Emulators take a small fraction of the computing time of simulators, and thus enable rigorous uncertainty assessment on even the most complex environmental models. Emulators are built by taking a sample of the simulator output and fitting a statistical model to that output. The mathematical techniques used to date are dominated by Gaussian processes, which postulate a smooth relationship between model inputs and its outputs (O'Hagan, 2006). For example, Reichert et al. (2011) presented a method founded upon linearized (simplified) model equations coupled to statistical descriptions of the emulation error, thereby allowing some of the model structure to be built into the emulator. However, the nonlinear aspects of aquatic biogeochemical model equations may not be effectively depicted by such Gaussian strategies, and thus we here employ Artificial Neural Networks (*ANN*s) due to their proven ability to reproduce complex relationships. *ANN*s are an example of biologically inspired optimization algorithms that usually provide high predictability (Haykin, 1994). We hypothesized that *ANN*s could offer a reliable emulator for exploring the posterior patterns of the Hamilton Harbour eutrophication model. The model structure of *ANN*s generally consists of one input layer, one or more hidden layers, and one output layer. In this study, we used a typical class of feed-forward neural network model called multilayer perceptrons (*MLP*s). The *MLP*s have been widely applied to solve a variety of complex environmental problems, demonstrating high predictive and forecasting power on plankton dynamics in limnological research (Jeong et al., 2006; Recknagel et al., 1998).

We developed three *ANN*s to predict ambient TP levels, chlorophyll *a* concentrations, and zooplankton carbon biomass. The three *ANN*s used nineteen (19) parameters from the eutrophication model. *TP* concentrations, total *TP* loading, river *TP* loading, chlorophyll *a* concentrations and zooplankton carbon biomass were additionally selected for predicting certain outputs. Posterior parameter vectors and the corresponding predictions of the eutrophication model form the basis for the training and test exercises (with a data partition ratio of 50:50) in a supervised manner. During the training process, genetic algorithms typically assisted the *ANN*s to reach an optimal structure with the best performance. Namely, a different structure (e.g., number of weight elements) in the hidden layer was used for each individual *ANN*. The genetic algorithms initialized 50 individual *ANN*s with different hidden layer structures, as candidates for the model with the best performance. We trained each *ANN* candidate in 1000 epochs to update the network weights and subsequently evaluated model fitness. This updating process was repeated 100 times. After the selection of the best model, sensitivity analysis was implemented. The sensitivity was quantified as the standard deviation of model outputs induced by input perturbations within the input mean ± standard deviation range (Principe et al., 2000). The above procedure was done by using the neural network software package, NeuroSolutions 5.0 (Lefebvre et al., 2005).

Results and discussion

Watershed modelling

To assess the performance of *SPARROW* model with time-varying parameters, we examined the plot of measured against predicted median values of the log-transformed total phosphorus loads (Fig. 3). The result showed good correspondence with the 1:1 line, with a slope of 0.99 and an r^2 value of 0.98. Comparing the phosphorus loads between Grindstone and Redhill Creek, the former was characterized by relatively higher median values ranging from 1.25 to 4.18 tons year^{-1}, whereas the model projections in the latter ranged from 0.97 to 3.42 tons year^{-1}. The uncertainty bounds of the model outputs varied between the two subwatersheds. The urbanized Redhill Creek watershed was characterized by wider credible intervals and considerable inter-annual variability of the standard deviation values (1.20–1.99 tons year^{-1}). On the other hand, the agriculturally dominated Grindstone Creek demonstrated lower standard deviation values with significantly narrower range (1.07–1.12 tons year^{-1}). The *TP* loadings followed similar inter-annual variability patterns in the two sub-watersheds, as the higher annual *TP* loading values in both sites were predicted in 1996 and 2008, while relatively low values were found in 1988, 1989, and 1999.

Generally, the means and standard deviations of the posterior parameter distributions suggest that substantial knowledge was gained for the seven *SPARROW* parameters after the Bayesian updating (Table 1). Most of the parameters were characterized by fairly distinct shifts of their central tendency relative to the prior assigned values, such as the land to water delivery coefficient (α), urban export coefficient (β_2), attenuation coefficients of small (ks_1) and large (ks_2) streams, whereas the agricultural export coefficient (β_1) and reservoir settling velocity (k_r) showed minor shifts of their posterior mean values. The posterior standard deviations were also significantly reduced relative to the prior values, and our results suggest that all parameters were well-identified and broadly in agreement with previous *SPARROW* applications (Alexander et al., 2002; Wellen et al., 2012, 2014). The model structural error (σ) had similar values with the error of the *SWALLOW* model, indicating that the current model configuration could still result in reasonable performance without the climatic forcing considered by Wellen et al. (2012).

We also investigated the inter-annual variability of the posterior values of the four time-varying parameters. The mean values of agricultural export coefficient (β_1) decreased from 0.255 in 1988 to 0.066 in 1999, and varied from 0.107 to 0.242 during the 2000s. By contrast, there was no systematic trend of the urban export coefficient (β_2), varying from 0.038 to 0.108. The mean posterior estimates of the attenuation coefficient of small streams (k_{s1}) were decreased from 1988 (0.231) to 2002 (0.082), and were subsequently increased up to 0.150 in 2009. The mean attenuation coefficient of large streams (k_{s2}) had a spike (0.046) in 1999, but it did not reach its highest level (0.120) derived for 1988. The relationships between the four time-varying parameters and the average annual stream flows measured at Redhill and Grindstone Creeks demonstrated interesting patterns (Fig. 4). There was a clear positive correlation between the agricultural export coefficient (β_1) and flow in both creeks, although there were a few profoundly outlying years associated with flow rates lower than 0.6 m^3 s^{-1}. As a result, only 14% and 29% of the variability of β_1 could be explained by the flows at Redhill and Grindstone Creeks, respectively. On the other hand, the flow rates of both creeks could account for a large portion of the variability of the urban export coefficient (β_2), i.e., 75% in Redhill Creek and 63% in Grindstone Creek. During periods of higher flow, the values of the attenuation coefficients for small (k_{s1}) and large streams (k_{s2}) were distinctly lower. The lower attenuation values during periods of higher flow are plausible and in agreement with previous theoretical

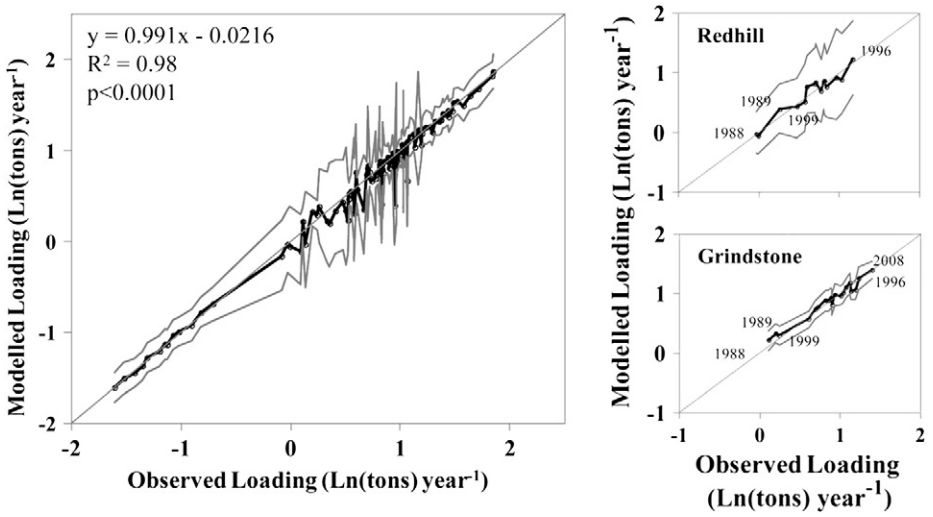

Fig. 3. Modelled and observed total phosphorus loading (kg day^{-1}) in Red Hill Creek (upper right panel), Grindstone Creek (lower right panel), and total loading to Hamilton Harbour (left panel) derived from the annual SPARROW model estimates. Black line corresponds to mean predicted values, while the gray lines correspond to the 95% credible intervals.

and empirical work on stream ecology, as the biotic (uptake) and abiotic (settling) processes responsible for attenuation have much less time to exert control on the nutrient load en route to the receiving water body when the streamflow rate is higher (Basu et al., 2011). On the other hand, longer hydraulic residence times typically imply a tighter coupling between the streambed and water column (Wood et al., 2007). Similar to Wellen et al.'s (2012) finding though, little of the year-to-year changes (<14%) of the attenuation estimates in small streams were explained by the annual flow rates, whereas a considerable amount of the variability (>60%) of the large stream attenuation coefficient can be attributed to the stream flow dynamics.

Downscaling annual SPARROW predictions to daily TP loads

We described the empirical relationship between the likelihood of low or high flow regimes during each seasonal period in each creek as a function of the daily precipitation. The logistic regression models in Redhill Creek were characterized by relatively greater slope coefficients (β) and lower intercept values (α) relative to Grindstone Creek. For example, in Redhill Creek, the slope β was 3.99 during July to September period, meaning that a unit of increase in the predictor variable *ln*(2-day average precipitation +1) results in a fourfold increase of the odds to shift into a high flow regime. In Grindstone Creek, the same coefficient was estimated to be 1.87. The intercept values (α) were −2.83 in Redhill Creek and −1.81 in Grindstone Creek or the probabilities to experience high flow conditions when the 2-day average precipitation tends to zero are approximately 4.5% and 14%, respectively. These results highlight the different hydrologic response of the two sites, as the relatively small size of the Redhill Creek subwatershed and the extensive urban development shape its flashier nature. Temporal variability of river flows and *TP* loads in Redhill and Grindstone Creeks are shown in ESM (Figure S1). Flows and *TP* loads show distinct seasonality patterns and the uncertainty bounds cover more than 90% of the observations. However, relative to the observed time series, the downscaled mean daily predictions failed to accurately reproduce the intra-annual variability in several years (e.g., 1993, 1999, 2007–2009), thereby introducing a summer overestimation and a winter underestimation of the flows and *TP* loads in both creeks.

Eutrophication modelling

The training exercise of the eutrophication model was intended to reproduce the "average" water quality patterns in the Harbour, as manifested over the past 22 years of the study period (1988–2010). The central tendencies and the underlying uncertainty of the nineteen parameter posteriors indicate that a substantial amount of knowledge was gained relative to the prior parameter specification (Table 2). For example, the detritus mineralization rate (φ), the zooplankton preference for detritus relative to phytoplankton (ω), the light extinction coefficient due to chlα (K_c) and the maximum phosphorus uptake rate for phytoplankton (P_{maxup}) demonstrated significant shifts in their posterior means relative to those specified prior to the calibration. Likewise, the posterior standard deviations of the detritus mineralization rate (φ), the zooplankton preference for detritus relative to phytoplankton (ω), the detritus sinking rate (ψ), and the fraction of refractory phosphorus buried into deeper sediment (β_P) were significantly narrower than the uncertainty of their priors. However, there were also several parameters that remained relatively unaltered with respect to posterior distributions, suggesting that we gained limited insights from the calibration dataset and/or the initial priors assigned were reasonable. Characteristic examples were the two parameters associated with the phytoplankton light limitation, namely, the half-saturation light intensity (I_s) and the background light extinction coefficient (K_b) as well as the phytoplankton sinking velocity (s). The delta index is an alternative measure applied to evaluate the degree of updating between parameter priors and posteriors (Fig. 5). The zooplankton preference for detritus relative to phytoplankton (ω), fraction of refractory phosphorus buried into deeper sediment (β_P), detritus mineralization rate (φ), and detritus sinking rate (ψ) had the greatest values of delta index (≥80%), indicating the highest distance between priors and posteriors. By contrast, the half-saturation light intensity (I_s), phytoplankton sinking loss rate (s) and background light extinction coefficient (K_b) were characterized by the lowest values (17%, 11%, and 3%, respectively), demonstrating minor shape changes between prior and posterior distributions.

The posterior medians along with the 95% credible intervals derived from the calibration of the eutrophication model closely followed the observed seasonal patterns for PO_4, *TP*, chl *a*, and total zooplankton biomass in the epilimnion (ESM Fig. S2). The model accurately reproduced the epilimnetic PO_4 levels, including the winter peak (≈14 μg L^{-1}) and summer low levels (≈1 μg L^{-1}). Our model also captured the two major peaks of the phytoplankton biomass (chl*a*) in May and August as well as the high zooplankton abundance in June. However, there was one month lag between the observed (August) and modelled (September) zooplankton biomass peak in the fall. In a similar manner, the model performance declined in the hypolimnion, as the PO_4

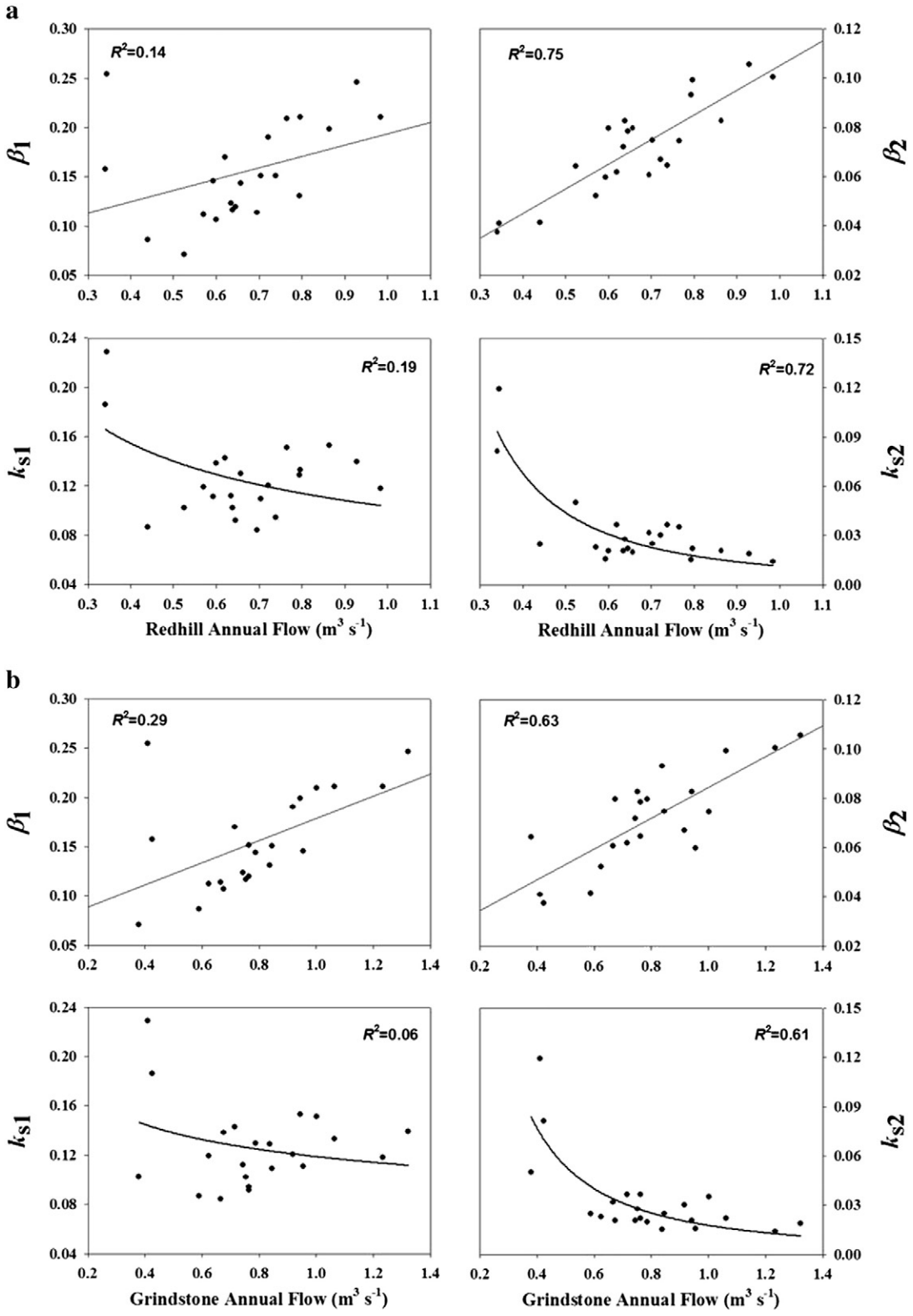

Fig. 4. Scatter plots of the posterior mean values of year-specific *SPARROW* model parameters (k_{S1}, k_{S2}, β_1, and β_2) against the average annual flows ($m^3 s^{-1}$) in Red Hill Creek (Panel A) and Grindstone Creek (Panel B). k_{S1} denotes the attenuation rate in first and second order streams (km^{-1}); k_{S2} denotes the attenuation rate in third and higher order streams (km^{-1}); β_1 represents the agriculture export coefficient (tons P km^{-2} $year^{-1}$); and β_2 represents the urban export coefficient (tons P km^{-2} $year^{-1}$).

concentrations were slightly over-predicted ($\approx$2–4 μg L^{-1}) in most months. The model also failed to reproduce the observed variation of the spring hypolimnetic *TP* concentrations and also underestimated by 6–10 μg TP L^{-1} the summer levels. The present results generally suggest that the structurally augmented eutrophication model (i.e., explicit consideration of the intra-cellular phosphorous storage and dynamic sediment P release) has improved its performance relative to the plankton model introduced by Ramin et al. (2012). The improved goodness-of-fit was particularly evident in regards to the epilimnetic phytoplankton and the zooplankton patterns during the fall season.

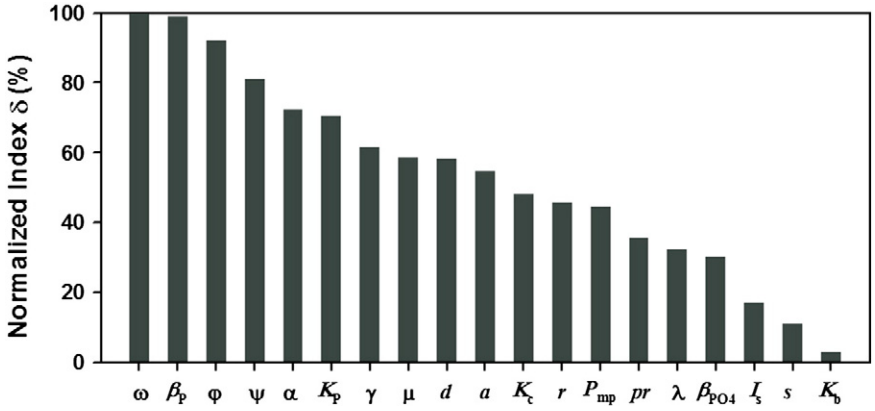

Fig. 5. Assessment of the changes in the shape of parameter distributions of the eutrophication model. The delta index is equal to zero if there is no difference between the two distributions, and equal to $\sqrt{2\log 2}$ if there is no overlap between the two distributions. All delta values are presented as percentages of this maximum value.

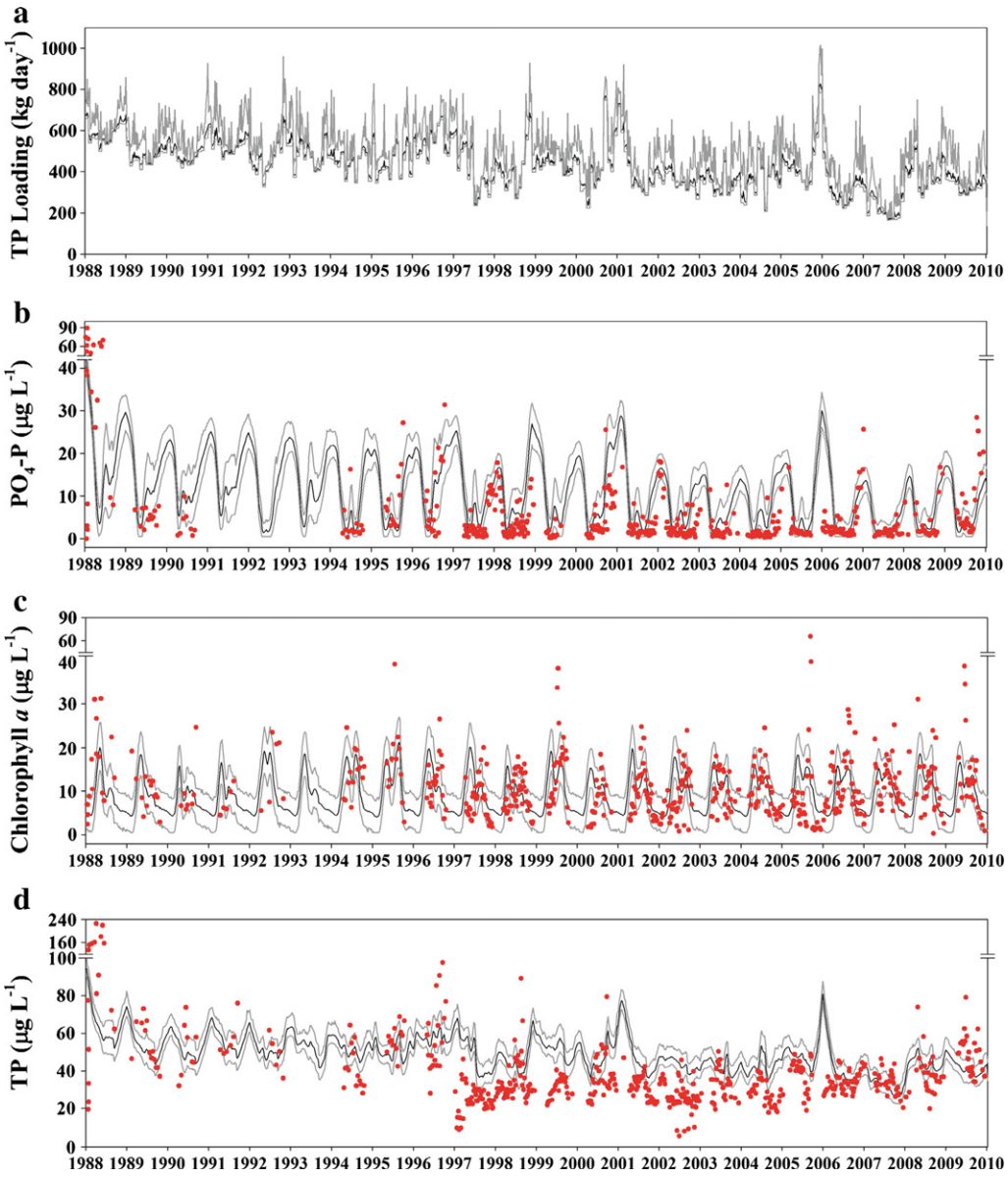

Fig. 6. Predictive comparison of the eutrophication model against the observed epilimnetic total phosphorus (*TP*), phosphate (PO_4), and chlorophyll a concentrations. Top panel shows the total phosphorus loading (kg day^{-1}) in the Hamilton Harbour during the 22-year validation period (1988–2010). Solid lines correspond to daily mean values while gray lines correspond to the 95% credible intervals.

Predictive confirmation of the integrated watershed-receiving waterbody model

The updated eutrophication model was then forced by the year-specific *SWALLOW* phosphorus loading predictions, transformed to daily inputs with the downscaling algorithm, to evaluate our predictive capacity over a 22-year period (1988–2009), where the total external *TP* loading declined from 600–800 kg day^{-1} in the late 1980s to 300–400 kg d^{-1} in the 2000s (Fig. 6a). This decreasing trend mainly stemmed from the point source control of the four wastewater treatment plants that discharge into Hamilton Harbour. In particular, the largest unit (Hamilton-Wentworth) began adding pickle liquor to the treatment process, successfully removing about half of total phosphorus from the discharge effluent (Charlton, 2001). The model was generally able to capture the inter- and intra-annual variability of the epilimnetic PO_4, *TP* and chlorophyll *a* concentrations with the majority of the observed data points being within the 95% credible intervals of model predictions. Simulated phosphate concentrations had good agreement with the typically observed seasonal pattern (mid-spring highs and summer lows) as well as the inter-annual variability, e.g., the relatively PO_4 high peak in 1996 and the fairly low spring levels in 2002 (Fig. 6b). The predicted mean chlorophyll *a* concentrations closely followed the inter-annual phytoplankton variability with respect to the magnitude and timing of the primary (end of spring) and secondary (late summer-early fall) blooms (Fig. 6c). However, we note that neither the mean predictions nor the uncertainty bounds could capture some extreme daily chlorophyll *a* values (>25 μg L^{-1}) occasionally observed towards the end of the summer period, e.g., 1995, 1999, 2005, 2006, and 2009. Even though the posterior simulations match closely the post-2005 levels, the *TP* concentrations from 1997 to 2004 appear to be overestimated (Fig. 6d). The latter discrepancies may partly stem from the inaccurate phosphorus loadings from the Woodward Wastewater Treatment Plant until the late 1990s (HHTT-CLR, 2004). In this study, efforts to develop an empirical formula that would correct the earlier loading estimates based on (the more reliable) recent data were unsuccessful. In addition, the fairly low *TP* levels in 1997 were likely the result of a prolonged and unusually high *Daphnia* abundance that profoundly controlled the algal biomass in the system, and consequently the amount of phosphorus that was sequestered in phytoplankton cells. The mechanisms for the unusual appearance of the large *Daphnia* population have not been unequivocally resolved although it was surmised that the abnormally low spring temperatures of that year caused a delay of fish migration into the system (including planktivores), which in turn released the zooplankton population from predation, resulting in zooplankton-mediated improvement of the water quality conditions (Lougheed and Chow-Fraser, 2001). Even if the latter hypothesis holds true, the current structure clearly does not have the capacity to accommodate this kind of top-down control, and thus our model is inherently inadequate to reproduce such a pattern. This is a characteristic example of Oreskes et al.'s (1994) assertion that it is practically impossible to completely confirm a model that was developed for an open natural system.

We subsequently applied principal component analysis (*PCA*) to elucidate the seasonal patterns of epilimnetic PO_4, *TP*, chlorophyll *a*, and zooplankton biomass. The basic rationale was that different phases of the intra-annual cycle may be regulated by separate processes and may therefore behave independently of each other. For this application of *PCA*, we formed four matrices (centered de-seasonalized data) of 12 columns (months of the year) and 22 rows (years of the study), and each row began with the month of the year when the serial correlation with the preceding month was the weakest (in the present case was between April–May). *PCA* was used to unravel the number of independent modes of variability, and the time of year in which they were most important (ESM Table S4). The selection of significant *PC*s was based on the Monte Carlo technique known as Rule *N* (Overland and Preisendorfer, 1982). Three eigenvalues accounted for 87% of the total variance for the PO_4 time series: the first *PC* (seasonal mode 1) represented the period from August to December; the second *PC* (seasonal mode 2) was characterized by high coefficients between January and April; and the third mode corresponded to the period of May and June, when the system usually experiences the lowest phosphate levels. Two *PC*s were selected for the monthly *TP* data, explaining about 84% of the total variance. The first seasonal mode coincided with the July-December period, while the second one represented the first half of the year (January-May). Four *PC*s accounted for 78% of the total variability in monthly chlorophyll *a* concentrations: the first seasonal mode was characterized by high positive coefficients during the summer stratified period (July-September) and negative coefficients during the winter (January and February). Interestingly, the phytoplankton biomass variability towards the end of the calendar year (November-December) corresponded to a distinct mode of behaviour. The third mode represented the two transitional periods of the system with respect to its hydrodynamic mixing regime; namely, the period spanning from the end of the ice cover until the stratification onset (March and April) and the period after the termination of the summer stratification (October). The fourth seasonal mode represented the period when the spring algal bloom typically occurs (May-June). The *PCA* for zooplankton biomass identified three eigenvalues that accounted for 80% of the total variance: similar to the first mode of the *TP* predicted monthly values, the first seasonal mode represented the period from July to December; the second mode corresponded to the January-February period; and the third mode represented the spring season (April-May) when the zooplankton abundance gradually increases. Notably, our model predicts that the highest zooplankton biomass levels typically occur in June, which was the month with no distinct signature in any of the extracted modes of variability.

Emulators of the eutrophication model posterior patterns

Multiple regression analysis was subsequently employed to examine the sensitivity of predicted summer *TP* and chlorophyll *a* concentrations to the model parameters (Manache and Melching, 2004). In all cases, the r^2 values were fairly high (>0.75) indicating that the relationship between input parameters and model outputs can be reasonably approximated as linear within the selected layout of the model posterior space (ESM Table S5). Based on the squared semi-partial correlation coefficient values, we identified the five most influential parameters underlying the *TP* and chlorophyll *a* predictions in each month from June to September. Generally, the summer *TP* variability was predominantly modulated by parameters associated with zooplankton abundance, such as the zooplankton assimilation efficiency (α), the fraction of zooplankton predation excreted as phosphorus (γ), the maximum zooplankton grazing rate (λ), and the zooplankton half-saturation constant for higher predation (*pred*). Notably, the detritus sinking rate (ψ) appears to have a consistently strong negative causal association with the summer *TP* epilimnetic levels. We also identified several influential parameters that are directly related to phytoplankton dynamics, e.g., the maximum phytoplankton growth rate (a), light extinction coefficient due to self shading effects (K_c), and the phytoplankton respiration rate (r) along with the sinking loss rate (s). The relative importance of those parameters on *TP* predictions varied among the different months; namely, the parameters associated with the zooplankton abundance were more influential in June and July, while the impact of parameters related to the phytoplankton characterization was more evident in August and September. Likewise, the chlorophyll *a* predictions were strongly influenced by two groups of parameters: *i*) the zooplankton-related parameters, e.g., assimilation efficiency (α), mortality rate (d), grazing half-saturation constant (μ), maximum grazing rate (λ), and half-saturation constant for fish predation (*pred*); and *ii*) the parameters associated with the characterization of the phytoplankton compartment, e.g., light extinction coefficient due to self shading effects (K_c), and respiration rate (r).

By relaxing the assumption of linearity, we subsequently examined the hypothesis that the *ANNs* could offer a more reliable emulator for exploring the model posterior patterns relative to multiple regression analysis. In a similar manner, we evaluated the influence of individual parameters and relevant variables of the eutrophication model (e.g., total and non-point phosphorus loading) on model endpoints of interest, such as epilimnetic *TP* levels, chlorophyll *a* concentrations, and zooplankton carbon biomass from June to September. The performance of the *ANNs* was better compared with the multiple linear regression models and all coefficients of determination (r^2) were greater than 0.90 (Fig. 7 & Table 3). First, we found that the *TP* concentrations in June and July were primarily affected by the zooplankton assimilation efficiency (α) and secondarily by the maximum zooplankton grazing rate (λ). On the other hand, the light extinction coefficient due to the self shading effects (K_c) and the maximum phytoplankton growth rate (a) were the most influential parameters in August and September. Second, the predicted chlorophyll *a* concentrations were primarily driven by the maximum zooplankton grazing rate (λ) in June and September, the zooplankton mortality rate (d) along with the maximum zooplankton grazing rate (λ) in July, and the phytoplankton respiration rate (r) in August. Finally, zooplankton carbon biomass was mostly sensitive to the light extinction coefficient due to the self shading effects (K_c) in June, July, and September, and the zooplankton assimilation efficiency (α) in August. Another parameter with significant impact on zooplankton abundance throughout the summer period was the maximum phytoplankton growth rate (a). These results were quite similar to those of the multiple regression analysis (Table S3), although the relative ranking of the various parameters differed between the two strategies considered to emulate the posterior space of the Hamilton Harbour eutrophication model.

Synthesis – next steps

We illustrated the development of a network of models that connect the watershed processes with the water quality dynamics of the Hamilton Harbour (Ontario, Canada). First, we opted for a parsimonious, data-driven strategy to characterize the average nutrient export rates from different land uses along with attenuation rates en route to the water body (*SPARROW*) and to subsequently accommodate the temporal variability of the nutrient loading (*SWALLOW*) over a 22-year study period. Our analysis suggests that the annual *TP* export from predominantly agricultural subwatersheds varied between 66–255 kg P m^{-2}, whereas the corresponding estimates from mainly urban areas ranged from 38–108 kg P m^{-2}. Based on the parameter posteriors, we can also infer that the proportion of *TP* attenuated per kilometer in small and large streams of the Hamilton Harbour watershed varied between 7.7–23.1% and 1.3–12.0%, respectively. One of the novel features of our framework was the introduction of a downscaling algorithm that transforms the annual phosphorus loading predictions to daily inputs for the eutrophication model. The proposed approach is founded upon the development of empirical relationships between the likelihood of low or high flow regimes during each season in each creek as a function of the daily precipitation. Our results presumably highlight the differences in the hydrologic response between areas characterized by agricultural activities and those with extensive urban development, although the size of the two catchments may be another confounding factor that shapes their behaviour. The downscaled daily flow predictions captured the observed variability in the two major tributaries (Redhill and Grindstone Creeks), but the corresponding *TP* loads tended to be underrepresented in the winter and somewhat overstated in the summer. These discrepancies primarily reflect the limitations of the

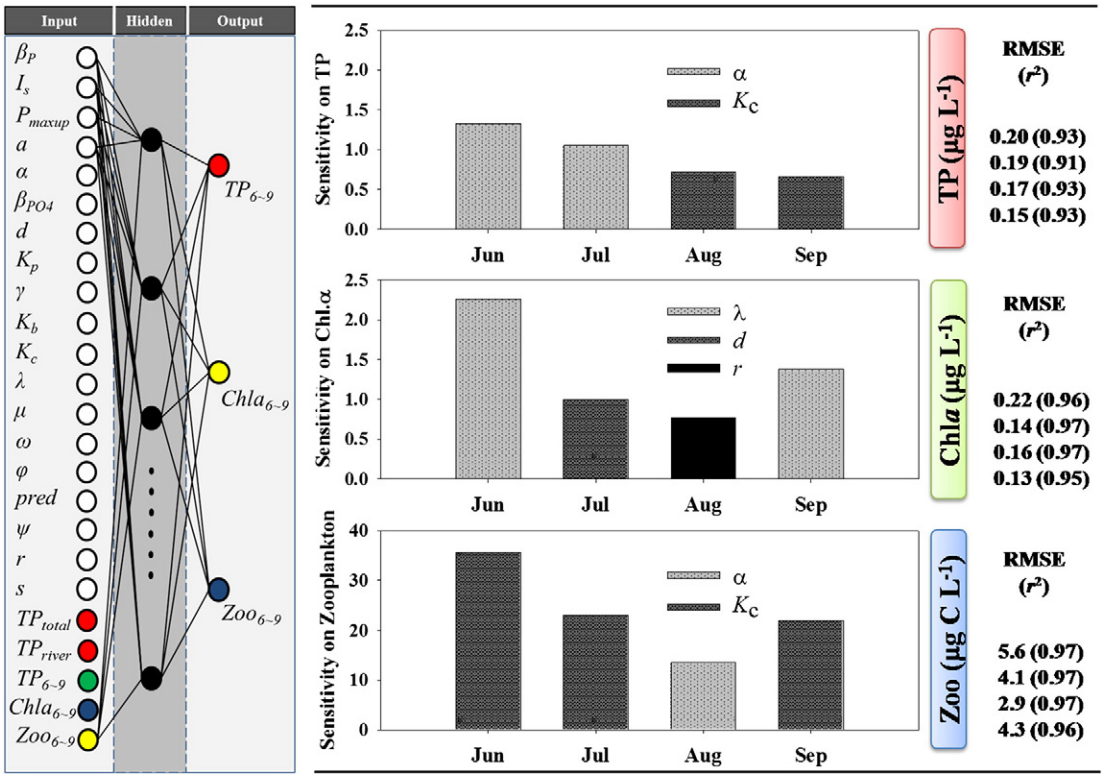

Fig. 7. Structural framework and predictive performance of the artificial neural networks (ANNs) built to emulate the eutrophication model posterior patterns. Input layer of the ANNs (in the left panel) includes all posterior parameters/variables of the eutrophication model. White-circled inputs were used for all the predictions of summer total phosphorus, chlorophyll *a*, and zooplankton (June to September), while colour-circled inputs were used only for the identical colour outputs; green-circled input ($TP_{6\sim9}$) was used for both chlorophyll *a* ($Chla_{6\sim9}$) and zooplankton ($Zoo_{6\sim9}$). The bar graphs depict standard deviations of the ANN's outputs, indicative of the impact of the most influential (in the middle) input parameters/variables. The right panels indicate model performances of the ANNs. The symbols of the eutrophication model parameters are provided in Table 2.

Table 3
Top five most influential parameters of *TP* concentration, chlorophyll *a*, and zooplankton summer monthly predictions, based on the determination coefficients (R^2) and sensitivity values derived from artificial neural networks. The sensitivity was quantified as the standard deviation of model outputs induced by input perturbations (i.e., mean ± standard deviation).

TP concentration				Chlorophyll α				Zooplankton			
Month	R^2	*Parameter*	*Sensitivity*	*Month*	R^2	*Parameter*	*Sensitivity*	*Month*	R^2	*Parameter*	*Sensitivity*
Jun	0.929	α	1.326	Jun	0.962	λ	2.262	Jun	0.969	K_c	35.613
		λ	1.207			α	1.925			a	30.068
		K_c	0.962			d	1.806			ω	18.019
		d	0.826			μ	1.545			r	17.101
		a	0.810			*pred*	0.553			d	15.286
Jul	0.910	α	1.051	Jul	0.973	d	0.994	Jul	0.968	K_c	23.013
		λ	0.815			λ	0.897			a	19.926
		K_c	0.806			μ	0.740			r	13.972
		a	0.727			α	0.537			ω	13.146
		d	0.704			ψ	0.352			d	9.829
Aug	0.925	K_c	0.717	Aug	0.967	r	0.774	Aug	0.970	α	13.620
		a	0.652			K_c	0.705			d	12.689
		α	0.630			λ	0.571			r	9.961
		ψ	0.599			a	0.548			K_c	9.550
		r	0.560			d	0.465			ω	8.659
Sep	0.926	K_c	0.661	Sep	0.954	λ	1.376	Sep	0.964	K_c	22.015
		α	0.630			d	1.216			a	17.998
		a	0.598			α	1.033			r	16.669
		ψ	0.532			μ	1.026			α	8.821
		r	0.495			K_c	0.543			ω	8.738

bivariate normal distributions used to draw daily *TP* concentrations from the annual average *TP* values predicted by *SPARROW*, while accounting for their potential codependence to the daily flows. The weak covariance between flows and concentrations in the study sites apparently poses constraints in the predictive capacity of our algorithm, although its Bayesian nature does allow the underlying error to be propagated and ultimately communicated through the uncertainty bounds of the eutrophication model predictions.

Counter to earlier work (Law et al., 2009), an implicit assumption of the present validation exercise was that the inter-annual variability of a natural ecosystem can be accommodated by temporally-constant rather than year-specific parameter characterization of the process-based model. Nonetheless, Law et al. (2009) demonstrated that several key planktonic processes can be characterized by considerable year-to-year variations, driven by weather variability, trends in the exogenous loading or other intrinsic ecosystem factors. In this regard, the discrepancy between predicted and observed water quality patterns in the highly eutrophic earlier years (late 1980s–early 1990s) of our simulation period is not surprising. Further, the systematic bias of the point-source loading during the 1990s may partly be responsible for the consistent overestimation of the ambient *TP* levels in the system. Similar to Hiriart-Baer et al.'s (2009) findings, our analysis showed that a number of water quality variables are characterized by regular seasonal patterns in the surface waters. Phosphorus demonstrated the lowest levels from the end of spring until the end summer and the highest values in the winter. Algal biomass, measured as chlorophyll *a*, was highest in the summer and lowest, albeit still elevated, in the winter. However, our model delineates a distinct seasonal mode of variability associated with the spring phytoplankton dynamics, which appears to deviate from the projections of recent empirical work in the Harbour (e.g., Fig. 3a in Hiriart-Baer et al., 2009). Namely, our analysis predicts a recurring spring phytoplankton bloom that can vary between 15–25 μg chl *a* L^{-1}, although the absence of an explicit ice cover submodel may introduce some uncertainty about the timing and magnitude of the spring phytoplankton bloom. It is important to note that the spring plankton dynamics in the Harbour are likely understudied, as the data collection typically starts after the first or second week of May which may coincide with the recession rather than the peak of the spring (diatom-dominated) bloom in Hamilton Harbour (Gudimov et al., 2010). While this piece of information may not be directly related to the summer water quality criteria and (ultimately) the delisting decisions of the system, it does allow to accurately quantifying the amount of biogenic material that deposits on the bottom of the system before the onset of stratification. This particulate pool may be easily decomposed during the summer and can potentially account for a significant proportion of the sediment oxygen demand (Gudimov et al., 2011).

The comparison between prior parameter distributions and their posterior patterns indicated that a substantial amount of knowledge was gained from the available data in regards to the characterization of the system. In particular, the posterior parameter specification dictates a strong dependence of zooplankton growth on phytoplankton grazing ($\omega = 0.035 \pm 0.018$), whereas their alternative food source (i.e., detritus) is subject to rapid sedimentation ($\psi = 0.478 \pm 0.028$ day^{-1}). Further, the relative contribution of the nutrient feedback loops, such as the release of phosphorus from the sediments ($\beta_P = 0.952 \pm 0.014$) or the bacterial-mediated nutrient recycling ($\gamma = 0.298 \pm 0.074$, $\varphi = 0.010 \pm 0.003$, $\beta_{PO4} = 0.352 \pm 0.052$), was fairly minimal, and therefore we can infer that the model postulates that the direct causal linkages among exogenous phosphorus loading → ambient phosphorus → phytoplankton → zooplankton predominantly shape the plankton patterns in the system. This conceptualization appears to deviate from the "web-like" structure introduced by earlier local modelling work (Gudimov et al., 2010, 2011; Ramin et al., 2011), but is on par with the findings of Hossain et al.'s (2012) Ecopath exercise, who asserted that the Harbour is a simple system with a linear food chain structure in the form of a "plant-herbivore-carnivore" sequence. In this regard, it is also not surprising that our attempts to emulate the model posterior space with both linear (multiple regression) and non-linear (artificial neural network) approaches were primarily based on parameters associated with the characterization of the zooplankton growth (e.g., assimilation efficiency, maximum grazing rate, half saturation grazing constant) and mortality/higher predation.

Consistent with the popular notion in the area, our calibration exercise downplays the relative contribution of the phosphorus fluxes from the sediments. For example, Mayer and Manning (1990) reported high phosphorus concentrations in solids collected from the sediment–water interface (>3500 mg kg^{-1}) and unusually high non-apatite inorganic phosphorus levels in sites adjacent to the municipal discharges, but they hypothesized that there is adequate ferric iron in the system to control the impact of the high P inputs. The same study also surmised that the retention of phosphorus in the sediments may be attributed to the ferric iron reduction, which subsequently results in the formation of an insoluble "Fe^{+2}-*other metal-P*" complex (Mayer and Manning, 1990). Nonetheless, Loh et al. (2013) reported month-long internal *P*

loading episodes in the Harbour, which can potentially lead to a significant hypolimnetic phosphorus accumulation. Recent empirical evidence also suggests that the phosphate concentrations in the hypolimnion can easily exceed the level of 50 μg PO_4 L^{-1} for extended period (3–4 weeks) during the late summer/early fall period (Environment Canada, unpublished data). Further, Gudimov et al. (2011) argued that a moderate increase of the sediment fluxes can significantly increase the number of violations of the water quality standard related to the epilimnetic *TP* concentration and could ultimately influence any future delisting decisions. Thus, the representation of the sediment diagenesis processes may be an important structural augmentation of the model, as the current simplified equation does not effectively account for the "memory" of the sediments when investigating the likelihood of the internal loading to exert control on the water quality conditions.

The pattern of hypolimnetic phosphorus accumulation also suggests that the summer epilimnetic environment may also be subjected to intermittent vertical nutrient intrusions, which in turn can have important ramifications on the abundance, composition or even predictability of phytoplankton dynamics (Jorgensen and Padisak, 1996; Soranno, 1997). Notably, our eutrophication model was not able to simulate the extreme algal biomass levels that are occasionally observed towards the end of the summer period and are usually associated with cyanobacteria dominance. Our limited ability to effectively reproduce the phytoplankton seasonal succession patterns and the cyanobacteria structural shifts was repeatedly identified as one of the knowledge gaps and outstanding challenges of the on-going restoration efforts in the Harbour (Gudimov et al., 2010; Ramin et al., 2012). Despite the long-term decrease of the *TP* levels, the Harbour experiences erratic outbreaks of noxious and (oftentimes) toxin-producing cyanobacteria (Murphy et al., 2003; Watson et al., 2008). This pattern seems to negate the basic premise of the existing *P* management paradigm that the capacity of cyanobacteria to outcompete the usual eukaryotic residents of the summer algal assemblage decreases under low phosphorus availability (Hyenstrand et al., 2001). Although our standpoint has been that there are more than one or two causal factors underlying the patterns of cyanobacteria dominance, we believe that Loh et al.'s (2013) empirical evidence of protracted summer episodes of internal *P* and *Fe* loading can be conceivably linked to the phosphorus-ferrous model (Molot et al., 2010). This model states that while phytoplankton productivity is controlled by *P*, significant diffusion of Fe^{+2} from anoxic sediments into waters near the euphotic zone is a prerequisite for cyanobacteria bloom formation (Molot et al., 2010). There are two compelling reasons why this hypothesis warrants further investigation: (*i*) even if the *RAP* nutrient loading reduction plans come into effect, the duration and severity of hypoxia will likely not improve significantly (Charlton, 2001), and thus the occurrence of internal loading events cannot be ruled out; and (*ii*) the phosphorus-ferrous model offers a reasonable explanation for the cyanobacteria outbreaks in mesotrophic (intermediate productivity) systems with TP concentrations below 20 μg L^{-1} that supposedly have low risk of bloom formation (Molot et al., 2010). The emphasis on *P* management has been successful and must remain the focus of the Hamilton Harbour restoration efforts. Our study suggests, however, that the on-going management plans should consider any plausible scenarios that can conceivably modulate the response of the system and may delay the establishment of the anticipated water quality conditions.

In conclusion, we presented a Bayesian framework that is uniquely suitable for developing integrated environmental modelling systems, as it can overcome the scale misalignment between ecological mechanisms of interest and available datasets, and may exploit diverse sources of information that differ with regards to the measurement error and resolution. Our network of models effectively connected the watershed processes with the phosphorus dynamics of the receiving waterbody in the Hamilton Harbour. The explicit consideration of the sediment-diagenesis processes and the incorporation of additional plausible mechanisms associated with the cyanobacteria dominance are two important directions, where additional structural complexity should be sought. The question arising is are we ready to mathematically depict or even to frame data collection efforts in this direction? Until we can give a positive answer to this question, we believe that the gradual incorporation of complexity, where possible and relevant, is the most prudent strategy. But any such model development should be tightly coupled with rigorous assessment of the underlying uncertainty and the Bayesian inference can be an invaluable ally in this frontier. Finally, we should not overlook Anderson's (2006) standpoint that prediction is not everything. We should not be afraid of complex models even if their structure reduces the predictive ability. Complex models offer excellent heuristic tools that allow insights into the direct, indirect, and synergistic effects of the numerous ecological mechanisms forming the foundation of system behaviour. They are an absolutely worthwhile scientific activity!

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. Aspects of the project were also supported by the Ontario Ministry of Research and Innovation (Early Researcher Award granted to George Arhonditsis). Dong-Kyun Kim received additional support by the Postdoctoral Fellowship Programme (2012-R1A6A3A03039357) from the Korea National Research Foundation. All the material pertinent to this analysis is available upon request from the corresponding author.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jglr.2014.04.005.

References

Alexander, R.B., Elliott, A.H., Shankar, U., McBride, G.B., 2002. Estimating the sources and transport of nutrients in the Waikato River Basin, New Zealand. Water Resour Res. 38 (12), 1268. http://dx.doi.org/10.1029/2001WR000878.

Ardenson, T.R., 2006. Confronting complexity: reply to Le Quere and Flynn. J. Plankton Res. 28, 877–878.

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., Brett, M.T., DeGasperi, C.L., Schindler, D.E., 2004a. Effects of climatic variability on the thermal properties of Lake Washington. Limnol. Oceanogr. 49, 256–270.

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

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.

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

Arhonditsis, G.B., Perhar, G., Zhang, W., Massos, E., Shi, M., Das, A., 2008b. Addressing equifinality and uncertainty in eutrophication models. Water Resour. Res. 44 (W01420), 15.

Arhonditsis, G.B., Stremilov, S., Gudimov, A., Ramin, M., Zhang, W., 2011. Integration of Bayesian Inference Techniques with Mathematical Modeling. In: Wolanski, E., McLusky, D.S. (Eds.), Treatise on Estuarine and Coastal Science, vol. 9. Academic Press, Waltham, pp. 173–192.

Basu, N.B., Rao, P.S.C., Thompson, S.E., Loukinova, N.V., Donner, S.D., Ye, S., Sivapalan, M., 2011. Spatiotemporal averaging of in-stream solute removal dynamics. Water Resour. Res. 47, W00J06. http://dx.doi.org/10.1029/2010WR010196.

Bayarri, M.J., Berger, J.O., Paulo, R., Sacks, J., Cafeo, J.A., Cavendish, J., Lin, C.-H., Tu, J., 2007. A Framework for Validation of Computer Models. Technometrics 49 (2), 138–154.

Beven, K.J., 2006. A manifesto for the equifinality thesis. J. Hydrol. 320, 18–36.

Beven, K.J., Alcock, R.E., 2012. Modelling everything everywhere: a newapproach to decision-making for water management under uncertainty. Freshw. Biol. 57, 124–132. http://dx.doi.org/10.1111/j.1365-2427.2011.02592.x.

Beven, K.J., Young, P.C., 2003. Comment on Bayesian recursive parameter estimation for hydrologic models by Thiemann, M., Trosset, M., Gupta, H. and Sorooshian, S. Water Resour. Res. 39 (5), 1116-1116. http://dx.doi.org/10.1029/2001WR001183.

Castelletti, A., Galelli, S., Ratto, M., Soncini-Sessa, R., Young, P.C., 2012. A general framework for Dynamic Emulation Modelling in environmental problems. Environ. Model Softw. 34, 5–18.

Charlton, M.N., 2001. The Hamilton Harbour remedial action plan: eutrophication. In. Verh. Int. Ver. Theor. Angew. Limnol. 27, 4069–4072.
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, 120pp.
Dietzel, A., Reichert, P., 2012. Calibration of computationally demanding and structurally uncertain models with an application to a lake water quality model. Environ. Model Softw. 38, 129–146. http://dx.doi.org/10.1016/j.envsoft.2012.05.007.
Edwards, A.M., Yool, A., 2000. The role of higher predation in plankton population models. J. Plankton Res. 22, 1085–1112.
Gilroy, E.J., Hirsch, R.M., Cohn, T.A., 1990. Mean square error of regression-based constituent transport estimates. Water Resour. Res. 26 (9), 2069–2077. http://dx.doi.org/10.1029/90WR00240.
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.
Hamilton Harbour RAP Technical Team (HHTT-CLR)., 2004. 1996–2002 Contaminant Loadings and Concentrations to Hamilton Harbour. 0-9733779-3-3.
Haykin, S., 1994. Neural Networks: A Comprehensive Foundation. Prentice Hall International, Inc., New Jersey (842 pp.).
Higdon, D., Kennedy, M., Cavendish, J.C., Cafeo, J.A., Ryne, R.D., 2004. Combining field data and computer simulations for calibration and prediction. SIAM J. Sci. Comput. 26, 448–466.
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 (2), 293–301.
Holling, C.S., 1959. The components of predation as revealed by a study of small-mammal predation of the European pine sawfly. Can. Entomol. 91, 293–320.
Hossain, M.M., Koops, M., Minns, C.K., Arhonditsis, G.B., 2012. Towards the development of an ecosystem model for the Hamilton Harbour, Ontario, Canada. J. Great Lakes Res. 38, 628–642.
Hyenstrand, P., Rydin, E., Gunnerhed, M., Linder, J., Blomqvist, P., 2001. Response of the cyanobacterium *Gleotrichia echinulata* to iron and boron additions – an experiment from Lake Erken. Freshw. Biol. 46, 735–741.
Jassby, A.D., 1999. Uncovering mechanisms of interannual variability from short ecological time series. In: Scow, K.M., Fogg, G.E., Hinton, D.E., Johnson, M.L. (Eds.), Integrated Assessment of Ecosystem Health. CRC Press, Boca Raton, FL, pp. 285–306.
Jeong, K.-S., Kim, D.-K., Joo, G.-J., 2006. River phytoplankton prediction model by Artificial Neural Network: Model performance and selection of input variables to predict time-series phytoplankton proliferations in a regulated river system. Ecol. Informl. 235–245.
Jorgensen, S.E., Padisak, J., 1996. Does the intermediate disturbance hypothesis comply with thermodynamics? Hydrobiologia 323, 9–21.
Kennedy, C.M., O'Hagan, A., 2001. Bayesian calibration of computer models. J. R. Stat. Soc. Ser. B 63, 425–464.
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.
Lefebvre, C., Fancourt, C., Principe, J., Gerstenberger, J., Samson, D., Euliano, N., Wooten, D., Lynn, G., Geniesse, G., Allen, M., Lucas, M., Marossero, D., 2005. NeuroSolution 5.0. NeuroDimension, Inc.
Loh, P.S., Molot, L.A., Nürnberg, G.K., Watson, S.B., Ginn, B., 2013. Evaluating relationships between phosphorus and iron release across three different water bodies. Inland Waters 3 (1), 105–118.
Lougheed, V.L., Chow-Fraser, P., 2001. Spatial variability in the response of lower trophic levels after biomanipulation in a freshwater marsh. J. Aquat. Ecosyst. Stress. Recover. 9, 21–34.
Manache, G., Melching, C.S., 2004. Sensitivity analysis of a water-quality model using Latin hypercube sampling. J. Water Resour. Plan. Manag. 130 (3), 232–242.
Mayer, T., Manning, P.G., 1990. Inorganic contaminants in suspended solids from Hamilton Harbour. J. Great Lakes Res. 16 (2), 299–318.
McMahon, G., Alexander, R.B., Qian, S.S., 2003. Support of total maximum daily load programs using spatially referenced regression models. J. Water Resour. Plan. Manag. 129, 315–329.
Molot, L.A., Li, G., Findlay, D.L., Watson, S.B., 2010. Iron-mediated suppression of bloom-forming cyanobacteria by oxine in a eutrophic lake. Freshw. Biol. 55, 1102–1117.
Murphy, T.P., Irvine, K., Guo, J., Davies, J., Murkin, H., Charlton, M.N., Watson, S.B., 2003. New microcystin concerns in the lower great lakes. Water Qual. Res. J. Can. 38, 127–140.
O'Hagan, A., 2006. Bayesian analysis of computer code outputs: a tutorial. Reliab. Eng. Syst. Saf. 91, 1290–1300.
Oreskes, N., Shrader-Frechette, K., Belitz, K., 1994. Verification, validation, and confirmation of numerical models in the earth-sciences. Science 263, 641–646.
Overland, J.E., Preisendorfer, R.W., 1982. A significance test for principal components applied to a cyclone climatology. Mon. Weather Rev. 110, 1–4.
Pappenberger, F., Beven, K.J., 2006. Ignorance is bliss: Or seven reasons not to use uncertainty analysis. Water Resour. Res. 42, W05302. http://dx.doi.org/10.1029/2005WR004820.
Principe, J.C., Euliano, N.R., Lefebvre, W.C., 2000. Neural and adaptive systems: Fundamentals through simulations. John Wiley & Sons, Inc., New York (656 pp.).
Qian, S.S., Stow, C.A., Borsuk, M.E., 2003. On Monte Carlo methods for Bayesian inference. Ecol. Model. 159, 269–277.
Qian, S.S., Reckhow, K.H., Zhai, J., McMahon, G., 2005. Nonlinear regression modeling of nutrient loads in streams: a Bayesian approach. Water Resour. Res. 41, W07012. http://dx.doi.org/10.1029/2005wr003986.
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. Model Softw. 26, 337–353.
Ramin, M., Labencki, T., Boyd, D., Trolle, D., Arhonditsis, G.B., 2012. A Bayesian synthesis of predictions from different models for setting water quality criteria. Ecol. Model. 242, 127–145.
Rao, Yerubandi R., Marvin, C.H., Zhao, J., 2009. Application of a numerical model for circulation, temperature and pollutant distribution in Hamilton Harbour. J. Great Lakes Res. 35, 61–73.
Recknagel, F., Fukushima, T., Hanazato, T., Takamura, N., Wilson, H., 1998. Modelling and prediction of phyto- and zooplankton dynamics in Lake Kasumigaura by artificial neural networks. Lakes Reserv. Res. Manag. 3, 123–133.
Reichert, P., Schuwirth, N., 2012. Linking statistical bias description to multi-objective model calibration. Water Resour. Res. 48, W09543. http://dx.doi.org/10.1029/2011WR011391.
Reichert, P., White, G., Bayarri, M.J., Pitman, E.B., 2011. Mechanism-based emulation of dynamic simulation models: Concept and application in hydrology. Comput. Stat. Data An. 55, 1638–1655.
Rode, M., Arhonditsis, G., Balin, D., Kebede, T., Krysanova, V., van Griensven, A., Van der Zee, S., 2010. New challenges in integrated water quality modelling. Hydrol. Process. 24, 3447–3461.
Runkel, R.L., Crawford, C.G., Cohn, T.A., 2004. Load Estimator (LOADEST): A FORTRAN Program for Estimating Constituent Loads in Streams and Rivers. Reston, Virginia, United States Geological Survey.
Sadraddini, S., Azim, M.E., Shimoda, Y., Bhavsar, S.P., Drouillard, K.G., Backus, S.M., Arhonditsis, G.B., 2011a. A Bayesian assessment of the PCB temporal trends in Lake Erie fish communities, J. Great Lakes Res. 37 (3), 507–520. http://dx.doi.org/10.1016/j.jglr.2011.06.005.
Sadraddini, S., Azim, M.E., Shimoda, Y., Bhavsar, S.P., Backus, S.M., Arhonditsis, G.B., 2011b. Temporal contaminant trends in Lake Erie fish: A dynamic linear modeling analysis. Ecotox. Environ. Safe. 74, 2203–2214. http://dx.doi.org/10.1016/j.physletb.2003.10.071.
Schwarz, G.E., Hoos, A.B.., Alexander, R.B., Smith, R.A., 2006. The SPARROW Surface Water-Quality Model – Theory. Applications and User Documentation. U.S. Geological Survey, Reston, Virginia.
Soranno, P.A., 1997. Factors affecting the timing of surface scums and epilimnetic blooms of blue green algae in a eutrophic lake. Can. J. Fish. Aquat. Sci. 54, 1965–1975.
Thiemann, M., Trosset, M., Gupta, H., Sorooshian, S., 2001. Bayesian recursive parameter estimation for hydrologic models, Water Resour. Res. 37, 2521–2535.
Vrugt, J.A., ter Braak, C.J.F., Clark, M.P., Hyman, J.M., Robinson, B.A., 2008. Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation. Water Resour. Res. 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. Can. J. Fish. Aquat. Sci. 65, 1779–1796.
Wellen, C., Arhonditsis, G.B., Labencki, T., Boyd, D., 2012. A Bayesian methodological framework for accommodating interannual variability of nutrient loading with the SPARROW model. Water Resour. Res. 48, W10505.
Wellen, C., Arhonditsis, G.B., Labencki, T., Boyd, D., 2014. Application of the SPARROW model in watersheds with limited information: A Bayesian assessment of the model uncertainty and the value of additional monitoring. Hydrol. Process. 28, 1260–1283.
Wood, P.J., Hannah, D.M., Sadler, J.P., 2007. Hydroecology and Ecohydrology: Past. Present and Future. John Wiley & Sons, Ltd., The Atrium (429 pp.).
Zhang, W., Arhonditsis, G.B., 2008. Predicting the frequency of water quality standard violations using Bayesian Calibration of Eutrophication models. J. Great Lakes Res. 34, 698–720.

TOWARDS THE DEVELOPMENT OF INTEGRATED MODELLING SYSTEMS IN AQUATIC BIOGEOCHEMISTRY: A BAYESIAN APPROACH

[Electronic Supporting Material]

Dong-Kyun Kim, Weitao Zhang, Veronique Hiriart-Baer, Christopher Wellen, Tanya Labencki, Duncan Boyd, George B. Arhonditsis*

* Corresponding author: e-mail: georgea@utsc.utoronto.ca, Tel.: +1 416 208 4858

List of Tables:

List of Figures:

Table S1: Mathematical description of the plankton model: equations and their constituent processes.

$\frac{dPO_{4i}}{dt} =$	$-P_{up\max}\frac{PO_{4i}}{Kp+PO_{4i}}\frac{P_{\max i}-P_{\text{int}\,i}}{P_{\max i}-P_{\min i}}PHYT_i$	Phytoplankton growth
	$+\frac{\beta_{PO4}\lambda\left((PHYT_i P_{\text{int}\,i})^2+\omega DET_i^2\right)}{\mu^2+(PHYT_i P_{\text{int}\,i})^2+\omega DET_i^2}fTz_i ZOOP_i P/C_{zoop}$	Zooplankton excretion
	$+\gamma d\, fTm_i\frac{ZOOP_i^3}{pred^2+ZOOP_i^2}P/C_{zoop}$	Zooplankton mortality/ higher predation
	$+\phi fT_i DET_i$	Mineralization
	$-kd(1-\sigma_t)(PO_{4,i}-PO_{4,i+1})\frac{2A_{i+1}}{V_i(H_{i+1}-H_{i-1})}$	Diffusion
	$+\alpha_{PO_4}PO4_{sed_i}e^{kt_{sed}(Temp_i-Tempsed_{ref})}\frac{A_i-A_{i+1}}{V_i}$	Sediment forcing
	$-(1\text{-}9.4\times(Fe+1400)^{-0.31})PO_{4i}$	Iron-induced precipitation
	$\pm ExchangesPO_{4Lake\ Ontario}$ †	Mass exchange with Lake Ontario
	$+PO_4 exog$ *	Exogenous loading
	$+Inflows\ PO_{4i}$ *	Inflows
	$-Outflows\ PO_{4i}$ *	Outflows
$\frac{dPHYT_i}{dt} =$	$a\frac{P_{\text{int}\,i}-P_{\min i}}{P_{\max i}-P_{\min i}}fT_i fL_i PHYT_i$	Phytoplankton growth
	$-r\, fT_i PHYT_i$	Phytoplankton respiration
	$-\frac{\lambda(PHYT_i P_{\text{int}\,i})^2}{\mu^2+(PHYT_i P_{\text{int}\,i})^2+\omega DET_i^2}fTz_i ZOOP_i$	Herbivory
	$-s(PHYT_i-PHYT_{i-1})\frac{A_i}{V_i}$	Settling
	$-kd(1-\sigma_t)(PHYT_i-PHYT_{i+1})\frac{2A_{i+1}}{V_i(H_{i+1}-H_{i-1})}$	Diffusion
	$\pm ExchangesPHYT_{Lake\ Ontario}$ †	Mass exchange with Lake Ontario
	$-Outflows\ PHYT_i$ *	Outflows
$\frac{dZOOP_i}{dt} =$	$\frac{\alpha\lambda\left((PHYT_i P_{\text{int}\,i})^2+\omega DET_i^2\right)}{\mu^2+(PHYT_i P_{\text{int}\,i})^2+\omega DET_i^2}fTz_i ZOOP_i$	Zooplankton grazing

	$-d\ fTm_i \frac{ZOOP_i^3}{pred^2 + ZOOP_i^2}$	Zooplankton mortality/ higher predation
	$\pm\ Exchanges ZOOP_{Lake\ Ontario}$ †	Mass exchange with Lake Ontario
	$-\ Outflows\ ZOOP_i$ *	Outflows
$\frac{dDET_i}{dt} =$	$r\ fT_i\ PHYT_i\ P/C_{phyto}$	Phytoplankton respiration
	$+\frac{\lambda[(1-\alpha-\beta)(PHYT_i\ P_{\text{int}\,i})^2 - (\alpha+\beta)\,\omega DET_i^{\,2}]}{\mu^2 + (PHYT_i\ P_{\text{int}\,i})^2 + \omega DET_i^{\,2}}\ fTz_i\ ZOOP_i\ P/C_{zoop}$	Detrivory minus zooplankton excretion
	$-\phi\ fT_i\ DET_i$	Mineralization
	$-\psi\ (DET_i - DET_{i-1})\frac{A_i}{V_i}$	Settling
	$+\ (OPsed_0)_i\ \sigma_t \frac{A_i - A_{i+1}}{V_i}$	Sediment forcing
	$-\ kd\ (1-\sigma_t)(DET_i - DET_{i+1})\frac{2A_{i+1}}{V_i(H_{i+1} - H_{i-1})}$	Diffusion
	$\pm\ Exchanges DET_{Lake\ Ontario}$ †	Mass exchange with Lake Ontario
	$+DETexog$ *	Exogenous loading
	$+\ Inflows\ DET_i$ *	Inflows
	$-\ Outflows\ DET_i$ *	Outflows
$\frac{dP\,\text{int}_{i,x}}{dt} =$	$P_{up\max}\frac{PO_{4i}}{Kp + PO_{4i}}\frac{P_{\max i} - P_{\text{int}\,i}}{P_{\max i} - P_{\min i}} - a\frac{P_{\text{int}\,i} - P_{\min i}}{P_{\max i} - P_{\min i}}\ fT_i\ fL_i P_{\text{int}\,i}$	Intracellular phosphorus content
$\frac{dPO_4 sed_i}{dt} =$	$\left(1-\beta_P\right)\frac{\left(sPHYT_i P_{\text{int}\,i} + \psi DET_i\right)V_i}{A_i - A_{i+1}}$	Sediment phosphorus deposition
	$-\ \alpha_{PO_4}\ PO_{4sed_i}\ e^{kt_{sed}(Temp_i - Tempsed_{ref})}$	Sediment phosphate release

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

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

$$fTm_i = e^{-ktm(Temp_i - Temp_{ref})}$$

$$fL_i = 2.718\,FD\frac{e^{-\alpha_{1i}} - e^{-\alpha_{0i}}}{ke_i H_i}$$

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

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

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

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

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

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

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

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

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

† The flow exchanges between Hamilton Harbour and Lake Ontario are based on the study by Klapwijk and Snodgrass (1985).

Table S2: Description of the parameters that were not considered during the Bayesian calibration of the eutrophication model.

Parameter	Description	Value and Unit
*kd**	Molecular plus eddy diffusion coefficient	$m^2 \cdot day^{-1}$
kt	Effect of temperature on phytoplankton processes	$0.005\ ^oC^{-2}$
ktz	Effect of temperature on zooplankton processes	$0.006\ ^oC^{-2}$
ktm	Effect of temperature on zooplankton higher predation processes	$0.069\ ^oC$
P/C_{zoop}	Phosphorus to carbon ratio for zooplankton	$0.029\ mg\ P \cdot (mg\ C)^{-1}$
$Temp_{ref}$	Reference temperature	$20\ ^oC$
ε	Shape parameter for the trigonometric function σ_t	0.85
α_{PO4}	Sediment phosphate release rate	$0.8\ day^{-1}$
kt_{sed}	Effects of temperature on sedimentation	0.004
$Tempsed_{ref}$	Sediment reference temperature	$20\ ^oC$

*The vertical diffusion is based on the study by Klapwijk and Snodgrass (1985).

Table S3: The statistical frameworks used for the training and predictive confirmation exercise of our integrated watershed-receiving water body model.

	Model training	*Model predictive confirmation*
SPARROW watershed model	$AL_i \sim N(Load_i, {\delta_i}^2)$ $Load_i \sim N(\mu_i, \sigma^2_{SPARROW})$ $\mu_i = \ln\left(\sum_{n=1}^{Ns}\sum_{j=1}^{J_i}\beta_n S_{n,j} e^{(-\alpha Z_j)} H^S_{i,j} H^R_{i,j}\right)$ $H^S_{i,j} = \prod_m \exp(-k_{s,m} L_{i,j,m})$ $H^R_{i,j} = \prod_l \exp(-k_r q_l^{-1})$ $\sigma^{-2}_{SPARROW} \sim gamma(0.001, 0.001)$ ${\delta_i}^2 = \sum_{x,y} Cov(DL_{i,x}, DL_{i,y})$	$AL_{i,t} \sim N(Load_{i,t}, {\delta_{i,t}}^2)$ $Load_{i,t} \sim N(\mu_{i,t}, \sigma^2_{SPARROW})$ $\mu_{i,t} = \ln\left(\sum_{n=1}^{Ns}\sum_{j=1}^{J_i}\beta_{n,t} S_{n,j} e^{(-\alpha Z_j)} H^S_{i,j,t} H^R_{i,j}\right)$ $H^S_{i,j,t} = \prod_m \exp(-k_{s,m,t} L_{i,j,m})$ $H^R_{i,j} = \prod_l \exp(-k_r q_l^{-1})$ $\sigma^{-2}_{SPARROW} \sim gamma(0.001, 0.001)$ ${\delta_{i,t}}^2 = \sum_{x,y} Cov(DL_{i,t,x}, DL_{i,t,y})$ $\theta_{t+1} \sim N(\theta_t, \varphi^2_{t+1}) \mid (\theta_{\min}, \theta_{\max})$ $\varphi^{-2}_{t+1} = 0.98^t \cdot \varphi_1^{-2}$ $\theta_1 \sim N(\theta_0, \varphi_1^2) \mid (\theta_{\min}, \theta_{\max})$ $\varphi_1^{-2} \sim gamma(a_\theta, \beta_\theta)$
Downscalling	$\log it[p_{FRd}] = \alpha_0 + \beta_0 X_d$ $X_d = \ln(precipitation_d + 1)$ $FR_d \sim Bernoulli(p_{FRd})$ $\kappa = FR_d + 1 = \begin{cases} 1 \text{ (low flow regime)} \\ 2 \text{ (high flow regime)} \end{cases}$ $\begin{pmatrix} Flow_{i,\kappa,d} \\ TP_{i,\kappa,d} \end{pmatrix} \sim MVN\left(\begin{pmatrix} \mu_{Flow_{i,\kappa,d}} \\ \mu_{TP_{i,\kappa,d}} \end{pmatrix}, \Omega_{i,\kappa}\right)$ $\Omega_{i,\kappa} = \begin{pmatrix} \sigma^2_{Flow_{i,\kappa}} & \rho_\kappa \sigma_{Flow_{i,\kappa}} \sigma_{TP_{i,\kappa}} \\ \rho_\kappa \sigma_{Flow_{i,\kappa}} \sigma_{TP_{i,\kappa}} & \sigma^2_{TP_{i,\kappa}} \end{pmatrix}$ $\overline{Load_{i,d}} = Flow_{i,k,d} \cdot TP_{i,k,d}$ $Load_{i,d} \sim N(\overline{Load_{i,d}}, \sigma^2_{SPARROW_{daily}})$	$\log it[p_{FRt,d}] = \alpha_0 + \beta_0 X_{t,d}$ $X_{t,d} = \ln(precipitation_{t,d} + 1)$ $FR_{t,d} \sim Bernoulli(p_{FRt,d})$ $\kappa = FR_{t,d} + 1 = \begin{cases} 1 \text{ (low flow regime)} \\ 2 \text{ (high flow regime)} \end{cases}$ $\begin{pmatrix} Flow_{i,\kappa,t,d} \\ TP_{i,\kappa,t,d} \end{pmatrix} \sim MVN\left(\begin{pmatrix} \mu_{Flow_{i,\kappa,t,d}} \\ \mu_{TP_{i,\kappa,t,d}} \end{pmatrix}, \Omega_{i,\kappa}\right)$ $\Omega_{i,\kappa} = \begin{pmatrix} \sigma^2_{Flow_{i,\kappa}} & \rho_\kappa \sigma_{Flow_{i,\kappa}} \sigma_{TP_{i,\kappa}} \\ \rho_\kappa \sigma_{Flow_{i,\kappa}} \sigma_{TP_{i,\kappa}} & \sigma^2_{TP_{i,\kappa}} \end{pmatrix}$ $\overline{Load_{i,t,d}} = Flow_{i,k,t,d} \cdot TP_{i,k,t,d}$ $Load_{i,t,d} \sim N(\overline{Load_{i,t,d}}, \sigma^2_{SPARROW_{daily}})$

Water quality model of receiving water-body	$WQ_d \sim N(WQ'_d, \sigma^2_{WQobs})$ $WQ'_d \sim N(f(\theta_1,\ \theta_2,\ x_d,\ y_0,\ \sum_i Load_{i,d}),\ \sigma^2_{WQ\,\mathrm{model}})$ $\ln(\theta_1) \sim N(\mu_\theta,\ \sigma^2_\theta)$ $\theta_2 \sim Beta(\alpha_\theta,\ \beta_\theta)$ $y_0 \sim MVN(\mu_{y_0}, \Sigma_{y_0})$ $\sigma^2_{WQobs}=(0.25 \cdot WQ_d)^2$ $\frac{1}{\sigma^2_{WQ\,\mathrm{model}}} \sim Gamma(0.001,0.001)$		$WQ'_{t,d} \sim N(f(\theta,\ x_d,\ y_0,\ \sum_i Load_{i,t,d}),\ \sigma^2_{WQ\,\mathrm{model}})$ $\ln(\theta) \sim MVN(\mu_\theta,\ \Sigma_\theta)$ $y_0 \sim MVN(\mu_{y_0}, \Sigma_{y_0})$ $\frac{1}{\sigma^2_{WQ\,\mathrm{model}}} \sim Gamma(a,b)$
Symbol	*Definition*	*Symbol*	*Definition*
i	The number of subwatershed/monitoring station	i	The number of subwatershed/monitoring station
j	The number of reaches in each subwatershed	j	The number of reaches in each subwatershed
m	The number of stream classes	m	The number of stream classes
n	The index of nutrient sources	n	The index of nutrient source
l	The number of lakes or reservoirs	l	The number of lakes or reservoirs
d	The index of day	t	The index of year
$N(.)$	Normal distribution	d	The index of day
AL_i	The natural logarithm of the measured annual load	$N(.)$	Normal distribution
$Load_i$	The latent "true" loading values at subwatershed *i*	$AL_{i,t}$	The natural logarithm of the measured annual load in year *t*
δ^2_i	The measurement loading error variance at subwatershed *i*	$Load_{i,t}$	The latent "true" loading values at subwatershed *i* in year *t*
μ_i	The natural logarithm of the annual load prediction at subwatershed i estimated by the SPARROW model	$\delta^2_{i,t}$	The measurement loading error variance at subwatershed *i* in year *t*
$\sigma^2_{SPARROW}$	The SPARROW model (process/structure) error variance	$\mu_{i,t}$	The natural logarithm prediction of the annual load at subwatershed i in year *t* estimated by the SPARROW model
β_n	The estimated export coefficient for source *n*	$\sigma^2_{SPARROW}$	The SPARROW model (process/structure) error variance
$S_{n,j}$	The nutrient mass from source n to reach *j*	$\beta_{n,t}$	The estimated source coefficient for source *n* in year *t*
α	Land to water delivery coefficient	$S_{n,j}$	The nutrient mass from source n to reach *j*
Z_j	Land surface characteristics associated with reach *j*	α	Land to water delivery coefficient
$H^S_{i,j}$	The fraction of nutrient mass originating in reach *j* and remaining at station *i* as a function of first order loss processes in streams	Z_j	Land surface characteristics data associated with reach *j*
$H^R_{i,j}$	The fraction of nutrient mass originating in reach *j* remaining at	$H^S_{i,j,t}$	The fraction of nutrient mass originating in reach *j*

	station i as a function of first order loss processes in lakes and reservoirs
$k_{s,m}$	The first order loss coefficient for stream class m (km^{-1})
$L_{i,j,m}$	The class m stream length (km) between reach j and station i
k_r	The first order loss coefficient or settling velocity ($m\ year^{-1}$)
q_l	The aerial hydraulic loading of the lake/reservoir ($m\ year^{-1}$)
$DL_{i,x}$	The loads on the arbitrary day x at station i calculated by the rating curve model
$DL_{i,y}$	The loads on the arbitrary day y at station i calculated by the rating curve model
$\text{log}it[.]$	Logistic regression model
α_0	The intercept term of the logistic regression models
β_0	The slope term of the logistic regression models
$Precipitation_d$	2-day moving average of the daily precipitation (logarithmic scale) used to predict the likelihood of the flow regime in day d
$Bernoulli(.)$	Bernoulli distribution
FR_d	The daily flow regime characterization
P_{FRd}	The likelihood of the daily flow regime
k	The index of the flow regime
$MVN(.)$	Multivariate normal distribution
$Flow_{i,k,d}$	The log-transformed daily flows for regime k in station i (only in Redhill and Grindstone Creek subwatersheds)
$TP_{i,k,d}$	The log-transformed daily TP concentrations for flow regime k in station i (only in Redhill and Grindstone Creek subwatersheds)
Ω_i	The covariance matrix between flow and TP concentrations
$\mu_{Flow_{i,k}}$	The mean assigned to each flow regime k after fitting the flow data in station i
	remaining at station i in year t as a function of first-order loss processes in streams
$H^R_{i,j}$	The fraction of nutrient mass originating in reach j remaining at station i as a function of first order loss processes in lakes and reservoirs
$k_{s,m,t}$	The first order loss coefficient for stream class m in year t (km^{-1})
$L_{i,j,m}$	The class m stream length (km) between reach j and station i
k_r	The first order loss coefficient or settling velocity ($m\ year^{-1}$)
q_l	The aerial hydraulic loading of the lake/reservoir ($m\ year^{-1}$)
$DL_{i,t,x}$	The loads on the arbitrary days x at station i calculated by the rating curve model
$DL_{i,t,y}$	The loads on the arbitrary days x at station i calculated by the rating curve model
$\theta_1,\ \theta_t,\ and\ \theta_{t+1}$	The time dependent SPARROW parameters at year 1, t, and $t+1$
$\theta_{min}\ and\ \theta_{max}$	The literature maximum and minimum values of the priors used with the time dependent *SPARROW* parameters
θ_0	The mean values of the literature-based priors used with the time dependent *SPARROW* parameters
$\varphi_1^2\ and\ \varphi_{t+1}^2$	The error variance of the time dependent *SPARROW* parameters for year 1 and $t+1$
$Gamma(.)$	Gamma distribution
$\alpha_\theta\ and\ \beta_\theta$	The shape parameters assigned to prior distribution of φ_1^2
$\text{log}it[.]$	Logistic regression models
α_0	The intercept term of the logistic regression models
β_0	The slope term of the logistic regression models
$Precipitation_{d,t}$	2-day moving average of the daily precipitation (logarithmic scale) used to predict the likelihood of the flow regime in year t and day d
$Bernoulli(.)$	Bernoulli distribution
$FR_{d,t}$	The daily flow regime characterization in year t

$\sigma_{Flow_{i,k}}^2$	The variance assigned to each flow regime k after fitting the flow data in station i	$P_{FRd,t}$	The likelihood of the daily flow regime in year t
$\mu_{TP_{i,k}}$	The annual average TP concentration calculated by dividing the SPARROW load prediction with the annual average flow in each creak in station i	κ	The index of the flow regime
$\sigma_{TP_{i,k}}^2$	The associated variance of the measured TP concentrations in station i	$MVN(.)$	Multivariate normal distribution
ρ_k	The correlation coefficient between flow and *TP* for the flow regime k	$Flow_{i,\kappa,t,d}$	The log-transformed daily flows for flow regime k in station i (only in Redhill and Grindstone Creek subwatersheds) in year t
$Load_{i,d}$	The daily non-point loading used to force the eutrophication model	$TP_{i,\kappa,t,d}$	The log-transformed daily *TP* concentrations for flow regime k in station i (only in Redhill and Grindstone Creek subwatersheds) in year t
$\sigma^2_{SPARROW_{daily}}$	The *SPARROW* process error expressed in daily load terms	$\Omega_{i,\kappa,}$	The covariance matrix between flow and TP concentrations for the flow regime k
WQ_d	The observation for water quality variables in receiving water-body	$\mu_{Flow_{i,k}}$	The mean assigned to each flow regime k after fitting the flow data in station i
WQ'_d	The latent "true value" used to parameterize the process-based models	$\sigma_{Flow_{i,k}}^2$	The variance assigned to each flow regime k after fitting the flow data in station i
σ^2_{WQobs}	The measurement error variance for water quality variables in receiving water-body	$\mu_{TPi,\kappa}$	The annual average TP concentration calculated by dividing the SPARROW load prediction with the annual average flow in each creak in station i
$f(\theta_1, \theta_2, x_d, y_0, \Sigma Load_{i,d})$	The prediction of water quality variables estimated by the process-based models	$\sigma_{TP_{i,\kappa}}^2$	The associated variance of the measured TP concentrations in station i
$\sigma^2_{WQmodel}$	The water quality model (process/structure) error variance	ρ_κ	The correlation coefficient between flow and *TP* for the flow regime k
θ_1	A time independent subset of the model parameters with log-normal distribution as prior	$Load_{i,t,d}$	The daily non-point loading used to force the eutrophication model in year t
μ_θ	The mean assigned to prior distribution of θ_1	$\sigma^2_{SPARROW_{daily}}$	The *SPARROW* process error expressed in daily load terms
σ_θ^2	The variance assigned to prior distribution of θ_1	$WQ'_{t,d}$	The latent predictive value of the process-based models in day d in year t
θ_2	A time independent subset of the model parameters with beta distribution as prior	$f(\theta, x_d, y_0, \Sigma Load_{i,t,d})$	The prediction of water quality variables estimated by the process-based models based on the joint posterior parameter patterns
α_θ *and* β_θ	The shape parameters assigned to prior distribution of θ_2	$\sigma^2_{WQmodel}$	The posterior of water quality model process/structure error variance
x_d	A time dependent set of the model forcing functions/boundary conditions	θ	The water quality model parameters

y_0	The model initial conditions	μ_θ	The posterior mean of water quality model parameters
μ_{y0}	The mean assigned to the multivariate normal distribution of y_0	Σ_θ	The posterior covariance matrix of water quality model parameters
Σ_{y0}	The covariance matrix assigned to the multivariate normal distribution of y_0	y_0	The model initial conditions
$Gamma(.)$	Gamma distribution	μ_{y0}	The mean assigned to the multivariate normal distribution of y_0
		Σ_{y0}	The covariance matrix assigned to the multivariate normal distribution of y_0
		a and b	The shape parameters of the posterior distribution of water quality model process/structure error

Table S4: Loading coefficients of the principal component analyses of the eutrophication model monthly outputs. Numbers in bold font denote coefficient values greater than 0.600.

Month	PO_4			TP		Chlorophyll *α*				Zooplankton		
	Factor 1	**Factor 2**	**Factor 3**	**Factor 1**	**Factor 2**	**Factor 1**	**Factor 2**	**Factor 3**	**Factor 4**	**Factor 1**	**Factor 2**	**Factor 3**
Jan	0.244	**0.911**	0.091	0.156	**0.873**	**-0.828**	0.123	-0.154	0.014	0.105	**0.962**	-0.010
Feb	0.270	**0.926**	0.152	0.242	**0.919**	**-0.769**	0.460	0.055	0.010	0.292	**0.931**	0.089
Mar	0.252	**0.916**	0.252	0.299	**0.927**	-0.202	0.480	**-0.704**	0.140	0.465	0.264	0.446
Apr	0.135	**0.877**	0.201	0.342	**0.882**	-0.180	0.036	**-0.909**	-0.049	0.297	-0.030	**0.908**
May	0.219	0.096	**0.897**	0.541	**0.767**	0.220	-0.091	-0.092	**0.818**	0.215	0.012	**0.946**
Jun	0.342	0.381	**0.796**	0.609	0.608	0.051	-0.095	0.195	**0.917**	0.503	0.305	0.510
Jul	0.514	0.488	0.520	**0.765**	0.456	**0.628**	0.145	0.016	0.421	**0.675**	0.292	0.389
Aug	**0.715**	0.349	0.425	**0.779**	0.346	**0.860**	0.262	0.058	0.226	**0.796**	0.252	0.291
Sep	**0.871**	0.232	0.255	**0.891**	0.220	**0.804**	0.294	0.327	0.146	**0.853**	0.254	0.245
Oct	**0.937**	0.192	0.180	**0.921**	0.247	-0.006	0.190	**0.807**	0.117	**0.920**	0.125	0.070
Nov	**0.916**	0.187	0.176	**0.883**	0.234	-0.174	**-0.764**	0.046	0.032	**0.871**	0.070	0.330
Dec	**0.862**	0.217	0.177	**0.853**	0.237	0.038	**-0.945**	-0.057	0.156	**0.720**	0.041	0.470
Eigenvalues	4.374	3.981	2.184	5.292	4.749	3.215	2.164	2.162	1.820	4.639	2.192	2.818
Total Variability	36%	33%	18%	44%	40%	27%	18%	18%	15%	39%	18%	23%

Table S5: Top five most influential parameters of *TP* and chlorophyll *a* summer monthly predictions, based on the standardized regression and the squared semi-partial correlation coefficient values derived from multiple regression analysis.

TP concentration					**Chlorophyll α**				
Month	***R^2***	***Parameter***	***β***	***R^2_{sp}***	***Month***	***R^2***	***Parameter***	***β***	***R^2_{sp}***
Jun	0.742	α	-1.788	0.158	Jun	0.852	λ	-2.369	0.192
		ψ	-0.875	0.150			α	-1.780	0.157
		γ	0.754	0.146			d	1.580	0.141
		K_c	-1.209	0.093			μ	1.709	0.138
		pred	-0.536	0.083			*pred*	-0.483	0.068
Jul	0.777	λ	1.175	0.354	Jul	0.884	d	2.077	0.245
		ψ	-1.060	0.220			μ	1.728	0.141
		α	-1.699	0.143			λ	-1.939	0.129
		K_c	-1.268	0.102			α	-1.440	0.103
		a	1.205	0.096			r	-0.524	0.055
Aug	0.777	γ	1.152	0.340	Aug	0.886	r	-1.211	0.292
		ψ	-1.316	0.340			d	1.597	0.145
		r	-1.199	0.286			K_c	-1.224	0.096
		s	-1.294	0.279			λ	-1.481	0.075
		K_c	-1.442	0.133			μ	1.194	0.067
Sep	0.820	ψ	-1.249	0.306	Sep	0.783	d	2.307	0.302
		γ	0.998	0.255			λ	-2.585	0.229
		r	-1.093	0.238			μ	2.129	0.214
		s	-0.962	0.154			*pred*	-0.817	0.194
		β_P	-0.787	0.153			α	-1.754	0.153

FIGURE LEGENDS

Figure S1: Observed versus downscaled daily flows ($m^3 s^{-1}$) and total phosphorus loading (kg day^{-1}) in Red Hill Creek (Panel A) and Grindstone Creek (Panel B), derived from the annual SPARROW model estimates. Black line corresponds to mean predicted values, while the gray lines correspond to the 95% credible intervals. The red dots represent the observed values in Red Hill Creek and Grindstone Creek during the 22-yr study period (1988-2010).

Figure S2: Bayesian calibration of the eutrophication model against the measured total phosphorus (*TP*), phosphate (PO_4), chlorophyll *a* concentrations, and zooplankton abundance during the seasonal cycle in the Hamilton Harbour. Solid lines represent the mean predicted values, while the dashed lines correspond to the 2.5^{th} and 97.5^{th} percentiles of the model predictions. The dots along with the error bars depict the observed mean values and the associated interannual variability during the 3-yr calibration period (2003-2006).

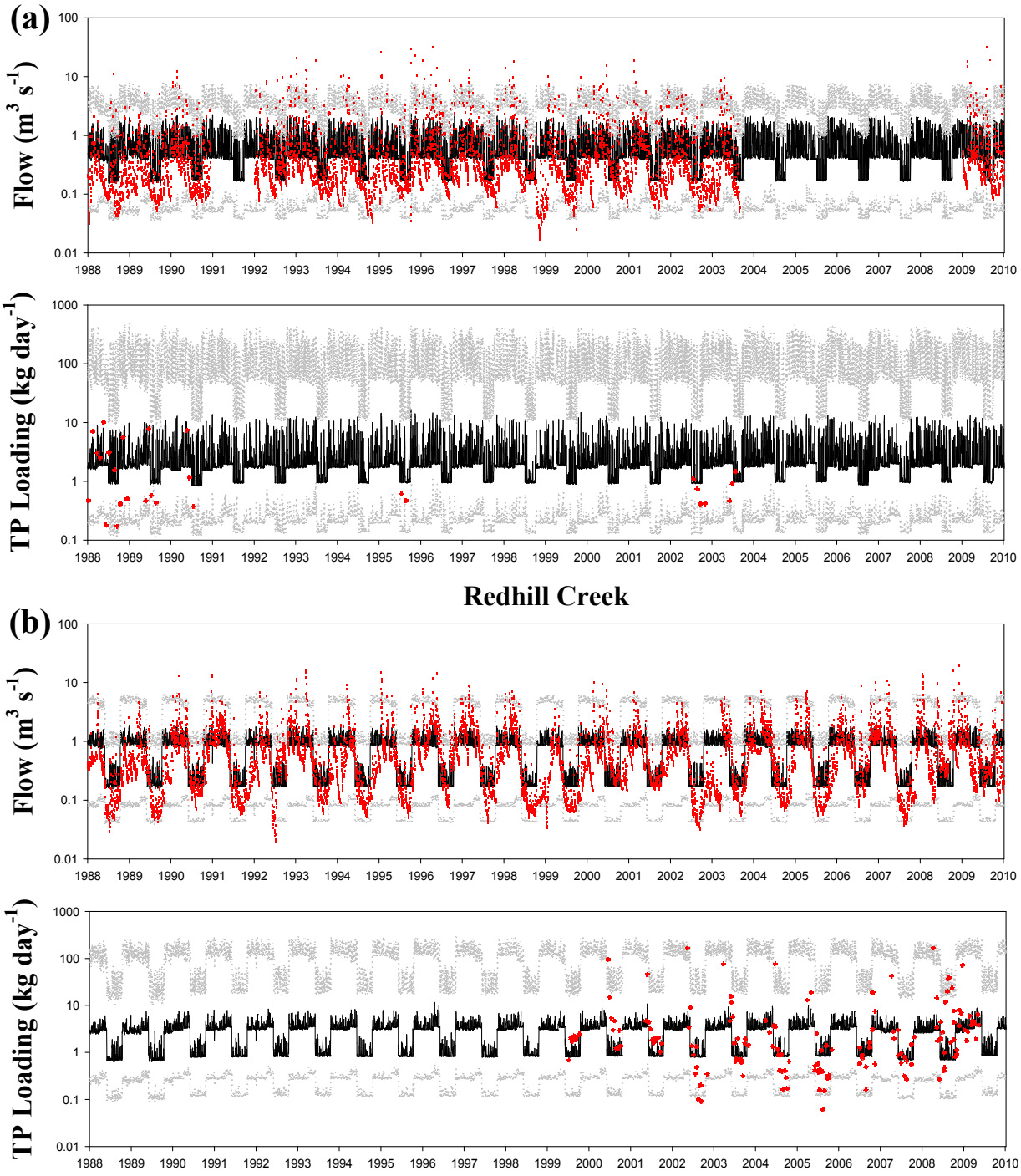
(a)
Flow (m3 s-1)
100
10
1
0.1
0.01
1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
TP Loading (kg day-1)
1000
100
10
1
0.1
1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
Redhill Creek
(b)
Flow (m3 s-1)
100
10
1
0.1
0.01
1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
TP Loading (kg day-1)
1000
100
10
1
0.1
0.01
1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
Grindstone Creek

Figure S1

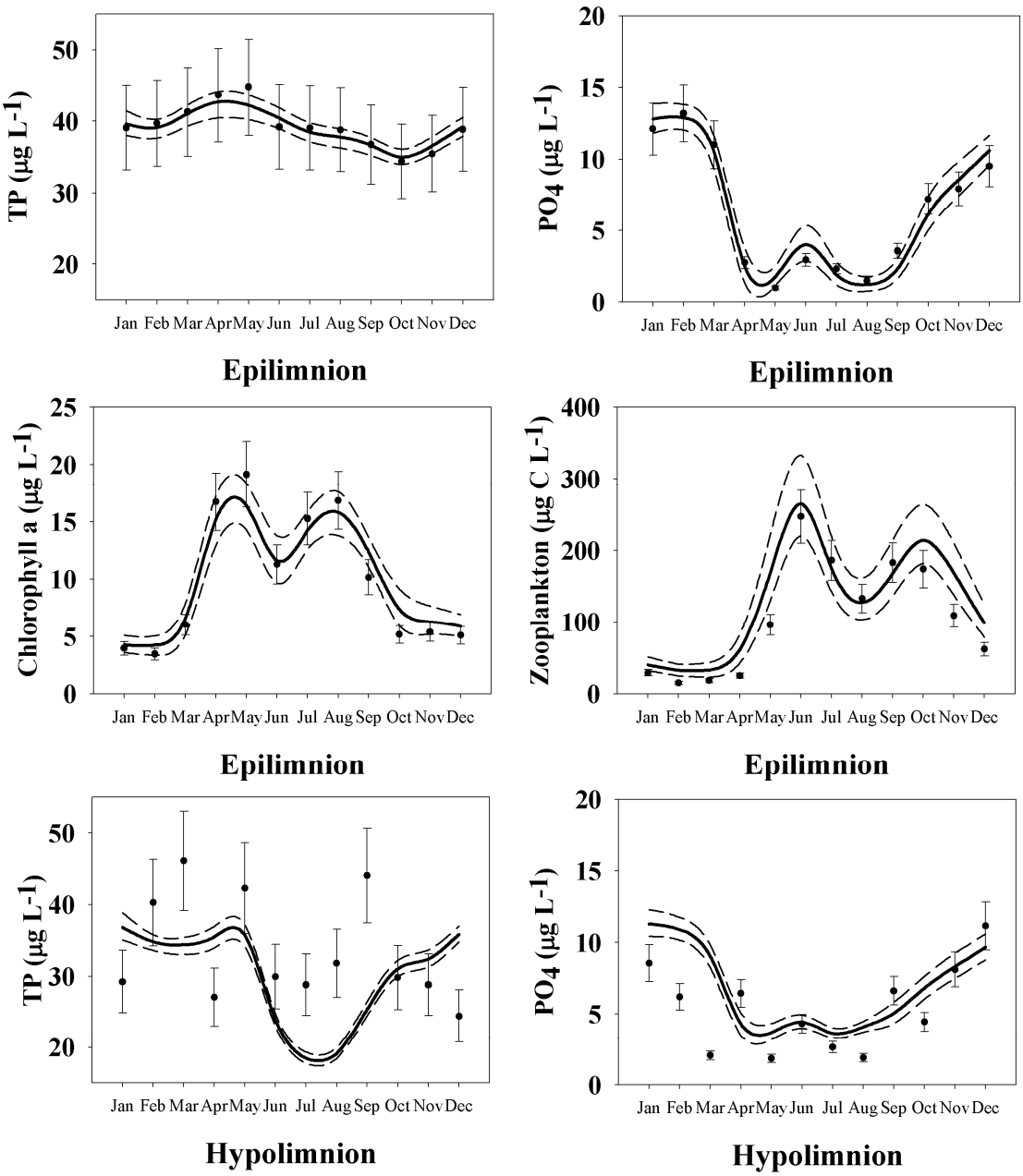
TP (µg L-1)
50
40
30
20
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Epilimnion
PO4 (µg L-1)
20
15
10
5
0
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Epilimnion
Chlorophyll a (µg L-1)
25
20
15
10
5
0
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Epilimnion
Zooplankton (µg C L-1)
400
300
200
100
0
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Epilimnion
TP (µg L-1)
50
40
30
20
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Hypolimnion
PO4 (µg L-1)
20
15
10
5
0
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Hypolimnion

Figure S2