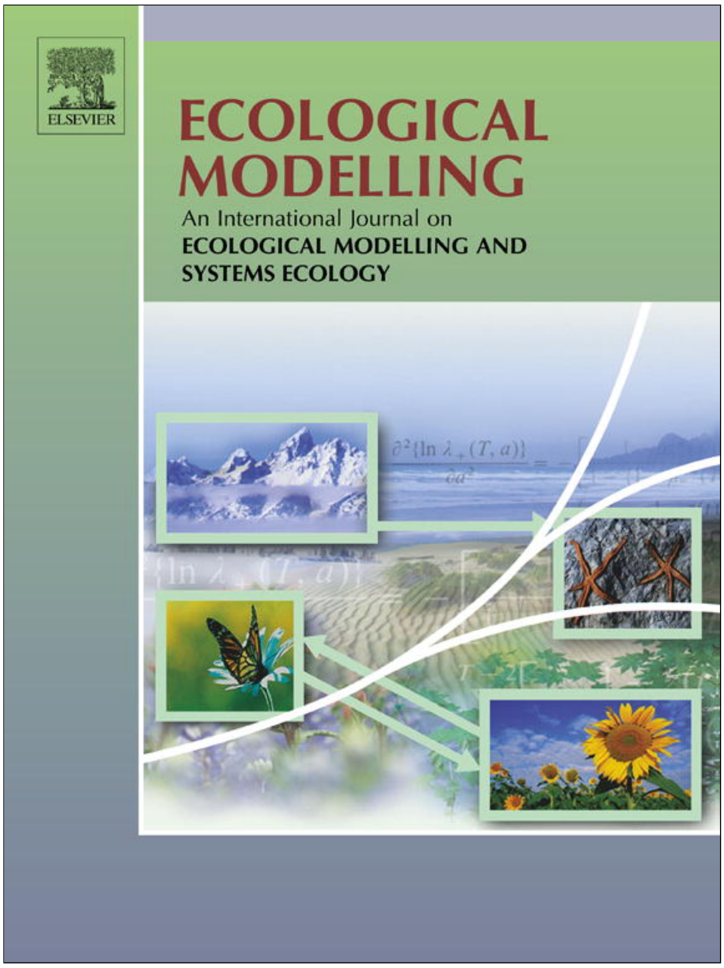
ELSEVIER
ECOLOGICAL MODELLING
An International Journal on
ECOLOGICAL MODELLING AND SYSTEMS ECOLOGY

Ecological Modelling 257 (2013) 66–79

Contents lists available at SciVerse ScienceDirect

Ecological Modelling

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

Can simple phosphorus mass balance models guide management decisions? A case study in the Bay of Quinte, Ontario, Canada

Weitao Zhang[a], Dong-Kyun Kim[a], Yerubandi R. Rao[b], Sue Watson[b], Shan Mugalingam[c], Tanya Labencki[d], Maria Dittrich[a], Andrew Morley[e], George B. Arhonditsis[a,*]

[a] Ecological Modelling Laboratory, Department of Physical & Environmental Sciences, University of Toronto, Toronto, Ontario, Canada M1C 1A4
[b] Environment Canada, National Water Research Institute, Burlington, Ontario, Canada L7R 4A6
[c] Lower Trent Conservation, Trenton, Ontario, Canada K8V 5P4
[d] Environmental Monitoring and Reporting Branch, Ontario Ministry of the Environment, Toronto, Ontario, Canada M9P 3V6
[e] Kingston Regional Office, Ontario Ministry of the Environment, Kingston, Ontario, Canada K7M 8S5

ARTICLE INFO

Article history:
Received 9 November 2012
Received in revised form 19 February 2013
Accepted 22 February 2013

Keywords:
Eutrophication
Phosphorus modelling
Bay of Quinte
Delisting objectives
Non-point loading
Environmental management

ABSTRACT

We test the capacity of an existing simple mass-balance total phosphorus (*TP*) model to evaluate nutrient loading scenarios in the Bay of Quinte, Ontario, Canada. Our study examines whether model parameters and loading inputs are well characterized and relevant to the current conditions in the Bay of Quinte and its drainage areas. We also identify critical data gaps and influential assumptions in regard to the uncertainty of model outputs and the credibility of predictive statements about the achievability of delisting objectives of the system. Our analysis shows that the model closely reproduced the observed variability of the *TP* seasonal averages during the calibration period 1972–2001, but its performance was significantly reduced when the actual predictive capacity was assessed in the 2002–2009 validation period. The most troublesome result is the inability of the model to reproduce the observed *TP* variability at temporal scales that are more meaningful from an environmental management point of view (i.e., monthly averages or daily snapshots from the system). Sensitivity analysis shows that several parameters associated with the role of the sediments were significant drivers of the model outputs, suggesting that considerable uncertainty exists in regard to the characterization of the sediments. The loadings from Trent River and the *TP* levels of the inflowing water masses from Lake Ontario predominantly shape the variability in the upper and lower segments of the Bay of Quinte, respectively. We also present a critical review of the suitability of the existing water quality criteria to depict the trophic status throughout the system. Our study contends that the summer average *TP* concentrations do not adequately reflect the prevailing conditions and that the development of proper water quality criteria should place more emphasis on inshore sites, where the eutrophication problems are more frequently manifested. Finally, we pinpoint factors unaccounted for by the original model that are likely to modulate the response of the system in its present state. We also discuss important directions of model structure augmentation and ways to optimize the spatial segmentation.

1. Introduction

The Bay of Quinte, a Z-shaped embayment at the northeastern end of Lake Ontario (Fig. 1), has a long history of eutrophication problems primarily manifested as frequent and spatially extensive algal blooms, predominance of toxic cyanobacteria, and hypolimnetic oxygen depletion. Legislation to reduce phosphorus in detergents along with upgrades at the local wastewater treatment plants resulted in substantial decline of point-source loadings during the 1970s, subsequently followed by a significant decrease of the ambient nutrient and phytoplankton biomass levels (Minns et al., 1986). Nonetheless, the Bay of Quinte was one of the 43 degraded sites around the Great Lakes designated by the International Joint Commission (*IJC*) as Areas of Concern (*AOC*s) in 1986 (Minns et al., 2011). Indicative of the continuing water quality and other persistent environmental problems, the Bay of Quinte was identified to be impaired in eleven out of the fourteen Beneficial Use Impairments (*BUI*s) (See also Glossary of Terms). Following the *AOC* designation, the Bay of Quinte Remedial Action Plan (*RAP*) was formulated through a wide variety of government, private sector, and community participants to provide the framework for actions aimed at restoring the system. In this regard, all the restoration efforts in the Bay were founded upon an "ecosystem" management approach, which was selected as a pragmatic means

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

0304-3800/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ecolmodel.2013.02.023

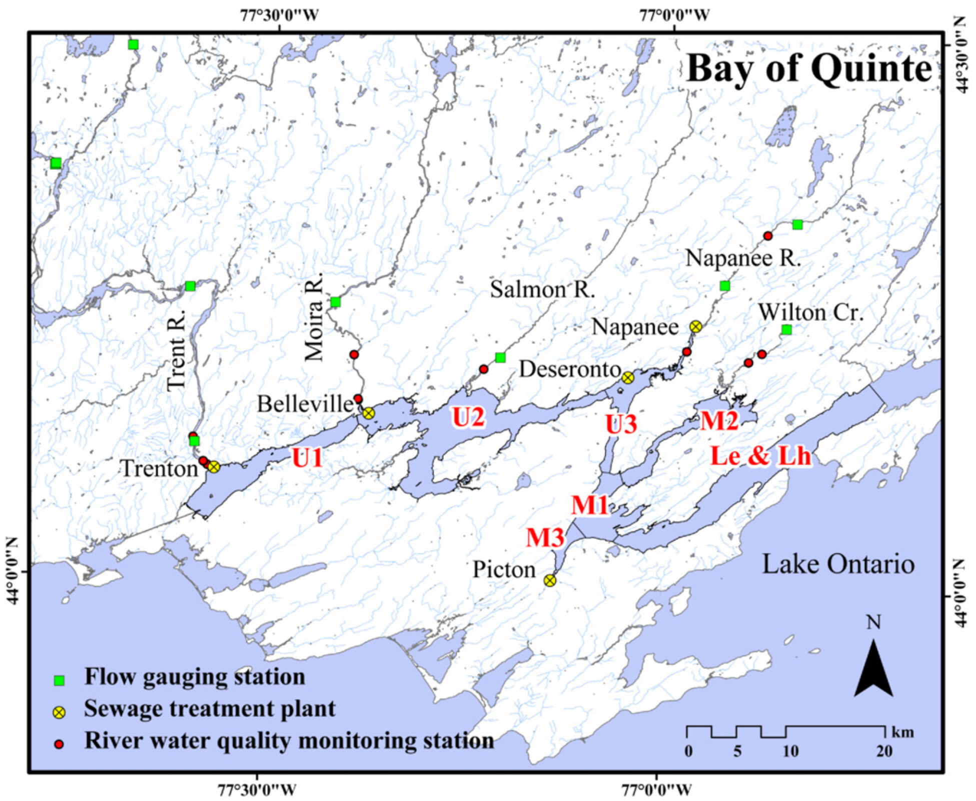

Fig. 1. Map of Bay of Quinte and segmentations of the original (PM_{2004}) and modified (PM_{2012}) model. The names in red represent the segment names for PM_{2012}. The Upper segment in PM_{2004} is separated into U_1, U_2 and U_3 in PM_{2012}. The Middle segment in PM_{2004} is separated into M_1, M_2 and M_3 in PM_{2012}. The Lower segment in PM_{2004} is vertically separated into L_e and L_h (L_e: Lower epilimnion; L_h: Lower hypolimnion) in PM_{2012}.

for accommodating the complexity pertaining to eutrophication control and addressing the combined effects of a suite of tightly intertwined stressors (Christie et al., 1986). The Bay of Quinte *RAP* process involves extensive monitoring work to provide evidence whether the designated beneficial uses have been restored and thus examine the likelihood of delisting the system as an *AOC* (Minns et al., 2011).

To this end, Minns et al. (2011) pointed out that the establishment of suitable environmental criteria should be one of the focal points of the contemporary efforts to delist the Bay of Quinte. The criteria selection process typically involves several critical steps, such as: (i) the identification of a measurable water quality variable that is a reliable predictor of the attainment of the beneficial use; (ii) the determination of the optimal numerical value of that water quality variable that allows distinguishing between impaired and non-impaired conditions (i.e., the so-called criterion level); (iii) the selection of a sensible resolution in time (e.g., annual/summer averages versus values from a number of snapshots from the system) and space (e.g., spatial averages throughout the system versus water quality trends of an offshore site) that will impartially depict the ecosystem state; and (iv) the evaluation of system compliance with the specified water quality criterion, using an operational procedure that is usually founded upon the collection of an adequate number of samples together with a rigorous statistical test and/or a properly validated process-based model (Borsuk et al., 2002; Zhang and Arhonditsis, 2008). It is also important that the natural system variability is explicitly accommodated, as the expectation of 100% attainment of the criterion level at all locations and at all times is probably unrealistic (Arhonditsis et al., 2007). The latter contention is particularly important when the water quality criteria setting process opts for a finer spatiotemporal resolution (Reckhow et al., 2005).

Striving for attainable water quality goals, the Bay of Quinte RAP (1993) identified three numerical objectives related to total phosphorus (*TP*), algal density, and submerged aquatic macrophytes as meaningful targets for significantly improving the Bay's water quality. The goals that emerged through a consensus on what were desirable and/or achievable targets were a summer average *TP* concentration of 30 $\mu g\,L^{-1}$, an average algal density in the range of 4–5 $mm^3\,L^{-1}$, and an increase in macrophyte abundance, so that 30% of the upper Bay has macrophyte coverage greater than 50% (Bay of Quinte RAP, 1993). Interestingly, whilst the likelihood of violations of these targeted values was not accommodated in a strict numerical sense, i.e., what the allowable exceedance frequency of the targets could be to consider the system in compliance, the original Stage 2 report, Time To Act, explicitly stated that these objectives were merely a "means to an end" and that there would still likely be periods of high algal densities, water clarity would still be more turbid than that of Lake Ontario, and that taste and odour problems in municipal and water supplies would still be present (Bay of Quinte RAP, 1993). Further, while an additional reduction of the exogenous nutrient loading seems to be the way forward, the determination of the critical levels is not a straightforward issue, as a series of perturbations associated with species invasions (dreissenid mussels, round goby) have significantly altered the ecosystem functioning and may compromise the outcome of the phosphorus management plans (Minns et al., 2011; Taraborelli et al., 2010). Given the social and economic implications of the associated policy decisions, we may have to contemplate whether the existing criteria adequately reflect the contemporary challenges of water quality management in the Bay of Quinte or if the type of probabilistic standards instructed by the U.S. EPA guidelines is more appropriate to accommodate the underlying uncertainty (Office of Water, 1997).

Environmental modelling has been an indispensable tool of the Bay of Quinte restoration efforts, where a variety of "*data-oriented*" and "*process-based*" models have been used for elucidating ecosystem dynamics and evaluating the potential consequences of alternative restoration actions (Diamond et al., 1996; Koops et al., 2006; Minns and Moore, 2004). In the context of nutrient loading management, Minns and Moore (2004) developed a simple mass-balance model, aiming to simulate phosphorus budgets in the Bay for the period 1972–2001 as well as to provide a method of forecasting future *TP* concentrations given various ecologically meaningful scenarios. Regarding the latter objective, one of the key projections of the Minns and Moore (2004) exercise was that the ominous forecasts of drier hydrology and higher summer water temperatures along with the presence of dreissenids may pose serious threat to the integrity of the system, even if the current exogenous *P* loading targets are achieved. The same study also predicted that the summer target of 30 μg *TP* L^{-1} for the upper Bay is unlikely to be met, as the mean tributary inflows lie above that value and the dreissenids accentuate the problem (Minns and Moore, 2004). While these predictions cast doubt on the system resilience and the likelihood to achieve further improvements, it is important to note that the Minns and Moore (2004) modelling study did not rigorously assess the effects of the uncertainty (structural uncertainty, parametric error, misspecified boundary conditions) underlying the model predictions on the projected system responses. Nor has it been discussed the suitability of simple input–output models to support predictive statements on finer temporal (monthly/daily) scales and thus accommodate percentile-based water quality standards (Zhang and Arhonditsis, 2008).

In this study, our primary objective is to examine the capacity of the existing simple mass-balance total phosphorus (*TP*) model to evaluate nutrient loading scenarios in the Bay of Quinte. In particular, we examine the soundness of the model parameterization and thus process characterization postulated by the Minns and Moore (2004) study. We also address several critical questions that are intended to shed light on the credibility of the original *TP* model, such as: Can the original model parameterization reproduce the observed short-term (e.g., monthly or day-to-day) variability in the system or is it more suitable for depicting variations of the summer *TP* average values? How efficient is the representation of the sediment processes and the role of the dreissenids in the original model, and how does the adopted approach compare with recent advances in the area of sediment diagenesis and benthic bioenergetic modelling? Finally, we undertake a critical review of the suitability of the existing criteria to depict the water quality status throughout the Bay of Quinte. Our study challenges the appropriateness of the summer average *TP* concentrations to adequately reflect the prevailing water quality conditions. We contend that the inference regarding the delisting of the system should be drawn by higher resolution information in time and that our water quality criteria should place more emphasis on inshore sites, where the eutrophication symptoms are more frequently manifested. It is also our hope that the lessons learned from the present exercise will serve as a showcase for the strengths/weaknesses of simple phosphorus models and their capacity to guide water quality management decisions.

2. Methods

2.1. Model description

Detailed description of the *TP* mass-balance model can be found in the original Minns and Moore (2004) study, and thus we just provide here the basic conceptual design along with the key model features. The spatial model segmentation consists of the upper, middle, and lower Bay sections (Fig. 1), coupled by water outflow from one section to the next. Each model section consists of two ordinary differential equations that describe: (a) the dynamics of phosphorus in the water column, determined by the inputs from the upstream section, tributary flows, atmosphere, sewage treatment plants and other point sources, refluxes from the bottom sediments, and phosphorus outputs via outflows to the next spatial compartment and settling to the bottom; (b) the variability of phosphorus in the sediment pool, driven by phosphorus inputs via settling from the water column and outputs via refluxes to the water column and losses to deeper sediment layers through burial (Fig. 2). At any given time, the volume of each spatial section is defined by a water budget, based on river inflows, inflows from the antecedent segment, precipitation, evaporation, and outflows to the subsequent segment. The resulting model was set up to run on a daily time-step over the 1972–2009 period, using the fourth-order Runge–Kutta method. In that time span, the period from 1972 to 2001 was considered to be the "historical" (or calibration) period and was parameterized with the same input data (tributaries, atmosphere, point sources, and precipitation) used by the Minns et al. (2004) budget analysis. The period from 2002 to 2009 was considered to be the "future" (or validation) period, in which we examined the actual predictive capacity of the model in the extrapolation domain. Additional inputs and outputs for the middle and lower segments of the Bay were also the backflows from Lake Ontario, based on the following scheme: (i) the backflow into the middle Bay is bringing water, having the *TP* concentration in the lower Bay, while an equal volume of water is displaced from the middle Bay with the *TP* levels in that section; (ii) an equal volume of water is in turn displaced from the lower Bay with the corresponding concentration, representing a lower Bay output; (iii) the lower Bay backflow brings water into the lower Bay having the *TP* concentration in Lake Ontario, and an equal volume of water is exported from the system with the concentration in that section.

The ordinary differential equation dealing with the sediments assumed a sediment mass defined by the product of the section accumulation area (defined in turn by the section area and accumulation extent) times a "sediment factor" (mass per unit area). The

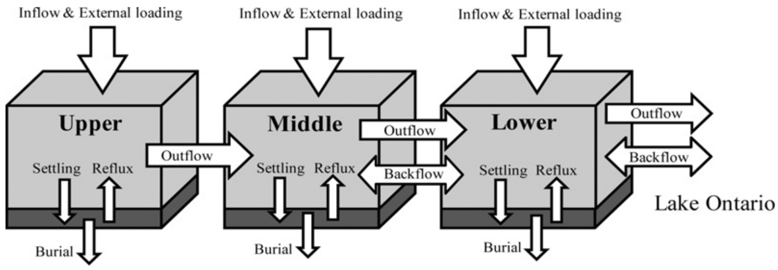

Fig. 2. Schematic representation of the total phosphorus balance of the Bay of Quinte model (PM_{2004}). The arrows indicate flows of mass through the system. The light grey boxes represent the water column and the dark grey boxes represent the sediment layer.

latter factor was assumed to be a function of sediment thickness, sediment water content, and sediment solids density. Phosphorus in the sediment pool was available for reflux to the water column and was also subject to burial. Phosphorus was added to the sediment pool via a settling process defined by a settling coefficient (day^{-1}) computed from the time-variant section mean depth (m) and a settling rate constant ($m\,day^{-1}$). The final process is the reflux from the sediment phosphorus to the water column. The representation of this process in the model was based on an exponential function, $R = a_{sed} \times e^{b_{sed} \times P_{sed}}$, fitted to monthly historical estimates of reflux rates R ($mg\,m^{-2}\,day^{-1}$), available from the original budget work of Minns et al. (2004), against the corresponding sediment P concentrations, P_{sed} ($mg\,g^{-1}$). The default values for a_{sed} and b_{sed} were 0.02 and 4.5, respectively. Further, Minns and Moore (2004) introduced an adjustment for the reflux rates to reproduce the trends observed in the post-1995 period, hypothesized to be associated with the invasion of dreissenid mussels. This "dreissenid mussel effect" was given a default value in the model of 0.01, simply by adding it to the a_{sed} coefficient in all years after 1995.

In the second phase of our analysis, the original model was modified by assigning temperature dependence to the reflux rates. The original spatial segmentation was also refined by splitting the upper Bay into three compartments (Fig. 1): (U_1) the segment that extends from the mouth of Trent River until the city of Belleville; (U_2) the segment that begins from the mouth of Moira River and comprises the Big Bay, Muscote Bay, and North Point Bay; and (U_3) the area influenced by the inflows of Napanee River, extending until the outlet of Hay Bay. In the middle Bay, we delineated three segments corresponding to the main stem (M_1) and the two adjacent embayments: Hay Bay (M_2), and Picton Bay (M_3). Finally, the lower segment of the Bay, representing the transitional area to Lake Ontario, was separated into the epilimnetic (L_e) and hypolimnetic (L_h) compartments.

The estimation of non-point phosphorus input from surrounding drainage basins closely follows the methods used by Minns et al. (1986, 2004). The rating curves (based on linear regression) provide daily estimates of tributary phosphorus loads from tributary flows. Daily flows were available throughout the simulation period for each of the five gauged tributaries (Trent River, Moira River, Salmon River, Napanee River, and Wilton Creek). The phosphorus concentration data for the rivers are based on the *PWQMN* (Provincial Water Quality Monitoring Network) dataset, collected by the Ontario Ministry of the Environment (*MOE*). Flows and *TP* loadings from un-gauged subwatersheds were accounted for using the Wilton Creek equivalents (*WCE*) approach, originally introduced by Minns et al. (1986). One *WCE* is the daily inflow from Wilton Creek which was then weighted by the ratio of the area of the un-gauged subwatershed to the Wilton Creek area (Minns et al., 2004). The un-gauged inflows and *TP* inputs into the upper Bay (U_1, U_2, and U_3) are 1.4, 4.7 and 0.1 *WCE*, respectively. The middle Bay (M_1, M_2, and M_3) receives 0.3, 1.8, 0.4 *WCE*, and the lower Bay is allotted 2.6 *WCE*. The monthly point source *TP* loads are based on effluent flows and *TP* concentrations from the Belleville, Trenton, *CFB* Trenton, Deseronto, Napanee, and Picton municipal sewage treatment plants (*STPs*). The effluents of Trenton and *CFB* Trenton *STPs* discharge into segment U_1, and Belleville discharges into segment U_2. Deseronto and Napanee *STPs* are located in the watershed of the U_3 segment. The only *STP* located in the middle/lower bay is the Picton treatment plant (M_3).

2.2. Model evaluation-sensitivity analysis

The evaluation of the original and modified models in the "historical" and "future" domains was based on both seasonal and month predictions against the corresponding observed values. Among the variety of model fit statistics available for evaluating model performance (Stow et al., 2003), we used the following summary statistics:

(i) root mean squared error (*RMSE*)

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(M_i - O_i)^2}{N}}$$

(ii) average error (*AE*)

$$AE = \frac{\sum_{i=1}^{N}(M_i - O_i)}{N} = \overline{M} - \overline{O}$$

(iii) relative error (*RE*)

$$RE = \frac{\sum_{i=1}^{N}|M_i - O_i|}{\sum_{i=1}^{N}O_i}$$

(iv) modelling efficiency (*ME*)

$$MEF = \frac{\sum_{i=1}^{N}(O_i - \overline{O})^2 - \sum_{i=1}^{N}(M_i - O_i)^2}{\sum_{i=1}^{N}(O_i - \overline{O})^2}$$

where N is the number of observations; O_i is the ith of N observations; M_i is ith of N predictions; and $\overline{O}$ and $\overline{M}$ represent the observation and prediction average values, respectively. The root mean squared error, average error, and relative error are all measures of the model prediction accuracy. Values near zero indicate a close match. The average error is a measure of aggregate model bias, though values near zero can be misleading because negative and positive discrepancies can cancel each other. The relative error is a scale-independent expression of the percentage discrepancy between predicted and observed values. The modelling efficiency measures how well a model predicts relative to the average of the observations. A value near one indicates a close match between observations and model predictions. A value of zero indicates that the model predicts individual observations as efficiently as the average of the observations does. Values less than zero indicate that the observation average would be a better predictor than the model outputs.

The next step of our study was focused on a detailed sensitivity analysis, aiming to evaluate the sensitivity of the model outputs to the parameter values and other assumptions made during its development. The quantification of these effects was achieved through independent perturbations on each element of the input vector and subsequent estimation of the corresponding changes of the predicted phosphorus concentrations at different segments. The procedure followed was based on Latin Hypercube sampling of a specified uncertainty zone of the model input space. In particular, each parameter was assigned a uniform distribution, representing a ±50% range of the final calibration value of the original/updated model (Table 1). The log transformed phosphorus daily loading from the five major rivers/creeks (Trent, Moira, Salmon, Napanee, Wilton) were assigned Gaussian distributions with mean and standard deviation values set equal to the rating curve mean estimates and standard error values, respectively.

3. Results

3.1. Model performance

Both the original and modified models closely matched the observed seasonal (May–October) *TP* patterns in the upper, middle,

Table 1
Parameter definitions and calibrated values of the original (PM_{2004}) model and modified (PM_{2012}) *TP* model. Reported ranges and types of distribution assigned to model parameters formed the basis for our sensitivity analysis.

Parameter	Unit	Calibrated Values		Min	Max	Distribution
		PM_{2004}	PM_{2012}			
Settling rate	m day^{-1}	0.113	0.113	0.0565	0.1695	Uniform
Sediment reflux a_{sed}	mg m^{-2} day^{-1}	0.02	0.02	0.01	0.03	Uniform
Sediment reflux b_{sed}	g mg^{-1}	4.5	4.5	2.25	6.75	Uniform
Dreissenid effect	g mg^{-1}	0.01	0.01	0.005	0.015	Uniform
Reflux temperature coefficient		–	1.06	–	–	–
Sediment solids density	g cm^{-3}	2.45	2.45	1.225	3.675	Uniform
Sediment water content	%	90	90	85	95	Uniform
Accumulation extent U_1, U_2, U_3	%	35	35	17.5	52.5	Uniform
Accumulation extent M_1, M_2, M_3	%	40	40	20	60	Uniform
Accumulation extent L_e	%	40	40	20	60	Uniform
Accumulation extent L_h	%	70	70	63	77	Uniform
Deposition constant	day^{-1}	1.3	0.3	0.15	0.45	Uniform
Sediment thickness U_1, U_2, U_3	cm	20	20	10	30	Uniform
Sediment thickness M_1, M_2, M_3	cm	15	15	7.5	22.5	Uniform
Sediment thickness L_e, L_h	cm	10	10	5	15	Uniform
Precipitation TP concentration	μg L^{-1}	15	15	7.5	22.5	Uniform
Wilton equivalents U_1		1.35	1.35	0.675	2.025	Uniform
Wilton equivalents U_2		4.69	4.69	2.345	7.035	Uniform
Wilton equivalents U_3		0.11	0.11	0.055	0.165	Uniform
Wilton equivalents M_1		0.32	0.32	0.16	0.48	Uniform
Wilton equivalents M_2		1.81	1.81	0.905	2.715	Uniform
Wilton equivalents M_3		0.35	0.35	0.175	0.525	Uniform
Wilton equivalents L_e		2.6	2.6	1.3	3.9	Uniform
Lake Ontario *TP*	μg L^{-1}	100%	100%	50%	150%	Uniform
Water back flow	m^3 day^{-1}	100%	100%	50%	150%	Uniform
Point loading U_1	kg day^{-1}	100%	100%	50%	150%	Uniform
Point loading U_2	kg day^{-1}	100%	100%	50%	150%	Uniform
Point loading U_3	kg day^{-1}	100%	100%	50%	150%	Uniform
Point loading M_3	kg day^{-1}	100%	100%	50%	150%	Uniform
Diffusion coefficient	m^2 day^{-1}	100%	100%	50%	150%	Uniform
Loading Moira River	kg day^{-1}					Log normal
Loading Napanee River	kg day^{-1}					Log normal
Loading Salmon River	kg day^{-1}					Log normal
Loading Trent River	kg day^{-1}					Log normal
Loading Wilton Creek	kg day^{-1}					Log normal

and lower segments of the Bay during the "historical" (or calibration) period (1972–2001) (left panels in Fig. 3). In particular, both models were able to reproduce the decreasing temporal trends in the upper segment, declining from about 80 μg L^{-1} in 1970s to approximately 40 μg L^{-1} towards the end of the calibration period. They also faithfully depicted the spatial gradients in the system, with distinctly higher *TP* levels in the upper segment relative to those experienced in the middle/lower Bay. The two models were subsequently forced with external inputs (point and non-point phosphorus loading, hydrological forcing, temperature) from the 2002–2009 period, while the rest of the parameterization (e.g., settling rates, sediment reflux coefficients) was set equal to the calibration values (Table 1). The actual predictive capacity of the two models was examined against the observed *TP* seasonal averages during the same period (Fig. 3). The results showed that the original (or PM_{2004}) model significantly underestimated the observed seasonal values in several years, e.g., the model predicts a seasonal *TP* average of 27.2 μg L^{-1} in the upper segment in 2005, whereas the corresponding observed value was 38.6 μg L^{-1}. Similar to the calibration results, the predictions of the updated (or PM_{2012}) model in the validation domain are distinctly higher than those obtained by the PM_{2004} model, e.g., in 2005, the seasonal *TP* value predicted by the modified model in the upper segment is 39.1 μg L^{-1}. Goodness-of-fit statistics reflect the relatively similar capacity of the two models to fit the seasonal *TP* levels in the Bay of Quinte during the calibration period, 1972–2001 (Table 2). Notably, the application of the original model was characterized by relatively higher R^2 (0.81), lower *AE* (1.54 μg L^{-1}), and higher *ME* (0.79) values in the upper segment relative to the corresponding fit statistics in the middle and lower segments. In the validation domain though, the PM_{2012} model appears to outperform in nearly all segments, suggesting that the reproduction of the *TP* concentrations – even in the fairly coarse seasonal scale – does improve by the explicit consideration of the role of water temperature in modulating the water column-sediment mass exchanges along with the revisit of the original model parameterization.

We further examined the capacity of the two models to provide accurate predictions with a finer temporal – monthly – resolution (right panels in Fig. 3). The original PM_{2004} model was clearly not able to capture the *TP* peaks typically experienced towards the late summer-early fall period in the upper Bay. The modelled range of the monthly *TP* concentrations was much narrower than the actual values and the predicted patterns profoundly failed to reproduce the substantial inter-annual variability in the system. Similar quantitative and qualitative predictions were also provided by the PM_{2004} model in the two downstream segments. On the other hand, the PM_{2012} model reproduced more closely the end-of-summer *TP* accumulation in the upper segment, while the predicted *TP* dynamics were more comparable with the observed year-to-year variability, e.g., the modelled monthly *TP* range was 11.5–52.0 μg L^{-1} during the validation period. Goodness-of-fit statistics confirmed the improved performance of the modified model relative to the original one when considering monthly instead of seasonally averaged predictions (Table 2). Namely, the R^2 values were 0.46 and 0.60 in the segments U_2 and U_3, respectively, while the corresponding value of the original model was approximately 0.09. Likewise, the *RMSE* values of the modified model were lower in U_2 (9.74 μg L^{-1})

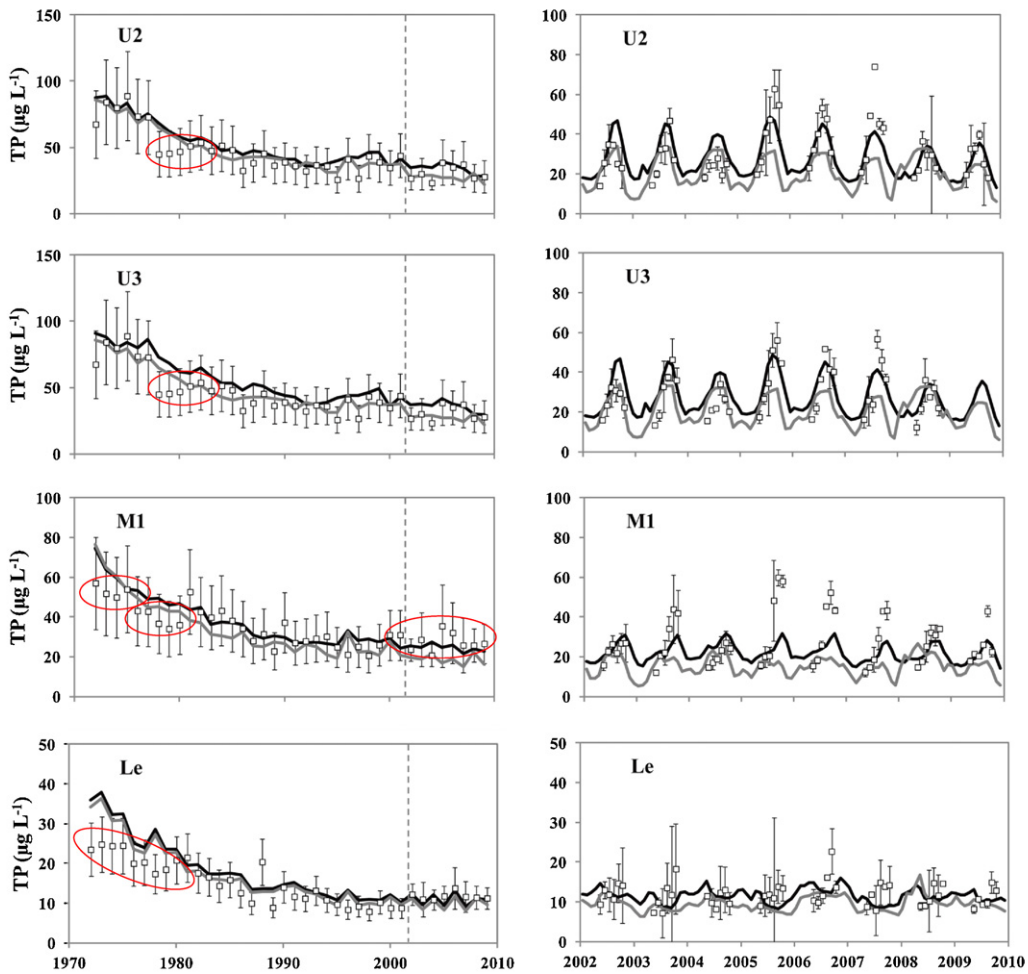

Fig. 3. Comparison between observed and predicted *TP* in Bay of Quinte. The left panel shows the results for the seasonal averages (May–October, 1972–2009) and the right panel shows the comparison for the monthly values (2002–2009). Grey lines represent the outputs of PM_{2004} and black lines represent the outputs of PM_{2012}. The square dots represent mean values of observed data and error bars denote the corresponding standard deviations. Grey dashed lines in the left panel illustrate the calibration (1972–2001) and validation (2002–2009) periods.

and U_3 (9.70 $\mu g\,L^{-1}$) relative to the corresponding error value of the original model (12.0 $\mu g\,L^{-1}$). Further, the marginally negative *ME* value (−0.04) suggests that the PM_{2004} model is as reliable predictor as the observed *TP* monthly mean in the upper segment. The *ME* values of the modified model were positive (>0.33) in the two spatial compartments of the upper Bay. In the middle segment, the performance of the PM_{2012} model slightly improved, but could not yet capture the late summer-early fall *TP* peaks. Aside from the increase of the R^2 value, we also note that the negative *ME* (−0.70) of the original model switched to a positive one (0.25) with the upgraded version. In the lower segment, the fit statistics suggest that the performance was not significantly improved by explicitly accounting for the dependence of the sediment phosphorus fluxes upon the water temperature variability. Not surprisingly, the corresponding

Table 2
Goodness-of-fit of the original (PM_{2004}) and modified (PM_{2012}) *TP* model in calibration (1972–2001) and validation (2002–2009) periods. Numbers in parentheses represent the validation results against the observed *TP* seasonal average concentrations during the 2002–2009 period.

Seasonal averages 1972–2001 (2002–2009)				Monthly averages 2002–2009			
PM_{2004}	Upper	Middle	Lower		Upper	Middle	Lower
R^2	0.81 (0.08)	0.68 (0.16)	0.77 (0.22)	R^2	0.09	0.03	0.02
RMSE	7.54 (7.52)	8.33 (10.55)	4.55 (2.25)	*RMSE*	12.0	16.0	4.12
AE	1.54 (−3.67)	−0.88 (−8.83)	2.37 (−1.84)	*AE*	−4.04	−9.23	−1.92
MEF	0.79 (0.74)	0.18 (−0.79)	0.33 (0.37)	*MEF*	−0.04	−0.70	−0.98
RE	12% (20%)	20% (33%)	21% (16%)	*RE*	28%	40%	26%

PM_{2012}	U_2	U_3	M_1	L_e		U_2	U_3	M_1	L_e
R^2	0.82 (0.26)	0.80 (0.40)	0.72 (0.02)	0.78 (0.32)	R^2	0.46	0.60	0.42	0.01
RMSE	8.82 (6.09)	11.9 (12.6)	7.45 (5.66)	5.30 (1.38)	*RMSE*	9.74	9.70	10.6	3.71
AE	5.23 (2.45)	8.83 (9.48)	2.92 (−2.82)	3.45 (−0.76)	*AE*	2.07	6.18	−3.23	−0.83
MEF	0.73 (0.82)	0.52 (0.65)	0.47 (0.48)	0.08 (0.76)	*MEF*	0.43	0.33	0.25	−0.61
RE	13% (14%)	19% (36%)	16% (18%)	27% (9%)	*RE*	23%	27%	27%	25%

RMSE – root mean squared error; *AE* – average error; *MEF* – model efficiency; *RE* – relative error.

ME values for the original and modified model were −0.98 and −0.61.

3.2. Sensitivity analysis

The sensitivity analysis of the modified model involved thirty one (31) parameters and twelve (12) external forcing functions, including five (5) non-point and four (4) point source loadings, and the water back flows along with the corresponding *TP* concentration from Lake Ontario (see Table 1). We used multiple regression analysis to examine the relative importance of the various model inputs to the predicted *TP* concentrations (Manache and Melching, 2004). Based on the squared semi-partial correlation coefficient values, we identified the top five (5) most influential parameters underlying the *TP* predictions in each segment of the Bay of Quinte (Table 3). Generally, the major drivers of the *TP* dynamics in the upper Bay (U_1, U_2, and U_3) are: (i) the non-point source loading from Trent River; (ii) the parameter b_{sed} associated with the sediment reflux rates; (iii) the parameters related to sediment characterization, e.g., sediment thickness, sediment water content, and sediment solid density; (iv) the deposition constant that directly determines the sediment burial rate. Among these factors, the external loading from Trent River is by far the most influential one. Interestingly, the squared semi-partial correlation coefficient value of the Trent River *TP* loading decreased from U_1 (74%) to U_3 (39%), accompanied by a gradual increase in the importance of internal loading (sediment reflux b_{sed}), i.e., 3%, 11% and 13% in U_1, U_2, and U_3, respectively.

In a similar manner to the upper Bay, the *TP* model predictions in the M_1 segment are strongly influenced by the Trent River loading (31%), the sediment reflux b_{sed} (12%), the water back flows (9%), as well as the parameters related to the sediment characterization (<6%). On the other hand, the spatial compartments M_2 and M_3 are relatively isolated, and thus the corresponding *TP* ambient levels are more sensitive to the sediment reflux rates than the external loading. In particular, the sediment reflux b_{sed} accounts for 27% and 19% of the total variability induced from our Monte Carlo experiments in M_2 and M_3, respectively. In the lower Bay of Quinte, *TP* dynamics are overwhelmingly influenced by the *TP* concentration of the inflowing water masses from Lake Ontario, e.g., squared semi-partial correlation coefficients are 73% and 92% in the epilimnion (L_e) and hypolimnion (L_h) of the lower segment. Our sensitivity analysis also shows that the actual hydraulic loading from Lake Ontario has a moderate impact on the ambient *TP* levels (L_e: 7% and L_h: 2%).

3.3. Phosphorus budget analysis

The phosphorus cycles as simulated by the original (PM_{2004}) and modified model (PM_{2012}) are illustrated in Fig. 4. The *TP* fluxes represent the total mass of phosphorus associated with the individual processes throughout the growing season (May–October) averaged over the 2002–2009 period. In the original model (Fig. 4a), the exogenous *TP* inputs from non-point, point, and aerial sources approximately contribute 243, 7, and 5 kg day^{-1} into the upper, middle and lower segments, respectively. The internal loading is also responsible for 311, 67, and 67 kg *TP* day^{-1} in the upper, middle and lower spatial compartments. Moreover, 331 and 227 kg day^{-1} of phosphorus are lost from the water column of the upper Bay through settling into the sediments and outflows into the middle segment. The inflowing water masses from Lake Ontario are an important vector of phosphorus transport (>1000 kg day^{-1}) into the lower segment, and subsequently into the middle Bay area (∼100 kg day^{-1}).

The refined spatial segmentation of the modified model (PM_{2012}) offers additional insights into role of the various *TP* fluxes in the system (Fig. 4b). In the U_1 segment, the external sources (phosphorus loading: 178 kg day^{-1}) and sinks (outflows: 221 kg day^{-1}) are greater than the two internal fluxes (155 and 120 kg day^{-1} for reflux and settling, respectively). On the other hand, the interplay between the internal phosphorus sources (sediment reflux: 377 kg day^{-1}) and sinks (particulate settling: 311 kg day^{-1}) appears to strongly modulate the *TP* dynamics in the U_2 segment relative to the exogenous loading (47 kg day^{-1}). The *TP* levels in U_3 and M_1 segments are dominated by the inflows from antecedent segments and the downstream outflows (∼300 kg day^{-1}), while the corresponding internal fluxes were significantly lower (60∼80 kg day^{-1}). The internal *P* loading contributes about 120 kg day^{-1} in the relatively isolated M_2 segment, whereas the external *P* inputs (5 kg day^{-1}) and the corresponding outflows (8 kg day^{-1}) are distinctly lower. In the lower area of the Bay of Quinte, the modified (PM_{2012}) model postulates a fairly closed loop, Lake Ontario → hypolimnion → epilimnion → Lake

Table 3
Model sensitivity analysis: Top five (5) most influential parameters of the *TP* predictions in each segment of the Bay of Quinte, based on the squared semi-partial correlation coefficient values derived from multiple regression analysis.

Rank	U_1		U_2		U_3	
1	Loading Trent River	74.3%	Loading Trent River	42.7%	Loading Trent River	38.5%
2	Sediment reflux b_{sed}	3.0%	Sediment reflux b_{sed}	11.1%	Sediment reflux b_{sed}	12.6%
3	Sediment thickness U_1	2.5%	Sediment solids density	5.4%	Sediment solids density	5.7%
4	Sediment solids density	2.3%	Sediment water content	5.4%	Sediment water content	5.6%
5	Sediment water content	2.2%	Deposition constant	4.1%	Deposition constant	4.3%
Rank	M_1		M_2		M_3	
1	Loading Trent River	30.6%	Sediment reflux b_{sed}	27.0%	Sediment reflux b_{sed}	18.5%
2	Sediment reflux b_{sed}	12.2%	Loading Wilton Creek	23.5%	Loading Trent River	15.9%
3	Water backflow	8.6%	Sediment Water Content	5.5%	Water backflow	8.9%
4	Sediment solids density	5.6%	Sediment thickness M_2	5.3%	Sediment solids density	5.6%
5	Sediment water content	5.5%	Sediment solids density	4.7%	Sediment water content	5.6%

Rank	L_e		L_h	
1	Lake Ontario *TP*	73.4%	Lake Ontario *TP*	91.9%
2	Water backflow	6.8%	Water backflow	2.0%
3	Loading Trent River	4.8%	Settling rate	0.3%
4	Sediment reflux b_{sed}	1.4%	Loading Trent River	0.2%
5	Sediment solids density	1.1%	Sediment solids density	0.1%

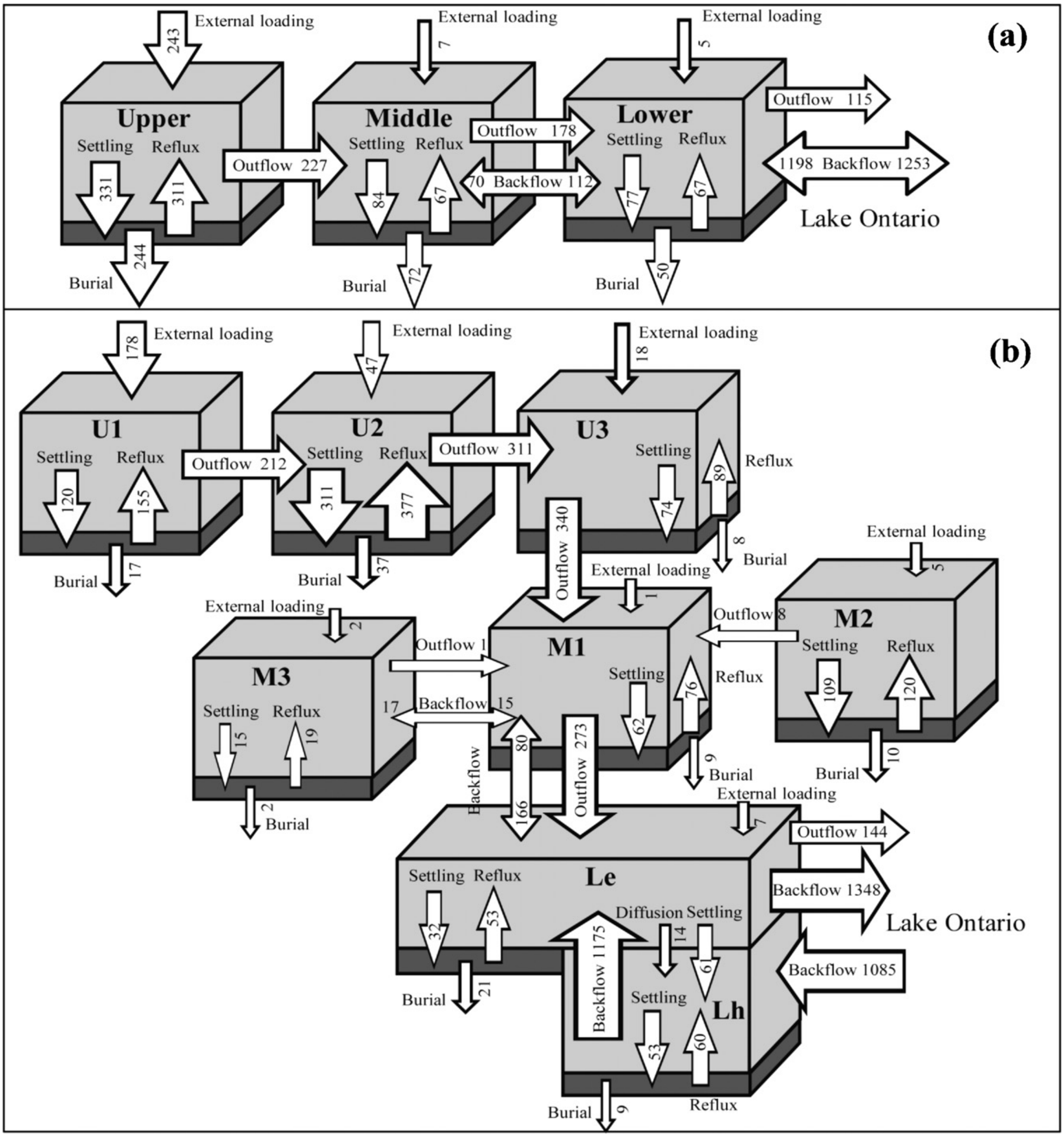

Fig. 4. Estimates of *TP* fluxes (kg day^{-1}) with (a) the original (PM_{2004}) and (b) the modified (PM_{2012}) *TP* model during the growing season (May–October) averaged over the 2002–2009 period.

Ontario, which represents an important pathway of phosphorus transport (>1000 kg day^{-1}). Another distinct feature of the updated model is the substantial reduction of the permanent loss of phosphorus mass from the system through the sediment burial. In the original model, the corresponding net phosphorus fluxes (sedimentation minus sediment reflux and burial) were −264 kg day^{-1} (upper segment), −89 kg day^{-1} (the middle segment), and −60 kg day^{-1} (the lower segment), which implies that the original sediment characterization was far from an equilibrium state. After revisiting the parameterization with the PM_{2012} version, the various sources and sinks of phosphorus in the sediments of the upper Bay are approximately balanced, i.e., 18 kg day^{-1} (U_1), 29 kg day^{-1} (U_2), and 7 kg day^{-1} (U_3). Similar conclusion could also be drawn by the phosphorus budgets in the rest of the segments, suggesting that the new calibration vector postulates that the sediments in the Bay of Quinte are close to an equilibrium state and thus do not act as a net sink of phosphorus.

3.4. Bayesian network of empirical models

Our study also examined the extent to which the predicted *TP* concentrations from our spatially explicit mass-balance model (PM_{2012}) can be used for forecasting other water quality variables of management interest (e.g., chlorophyll *a*). To address this

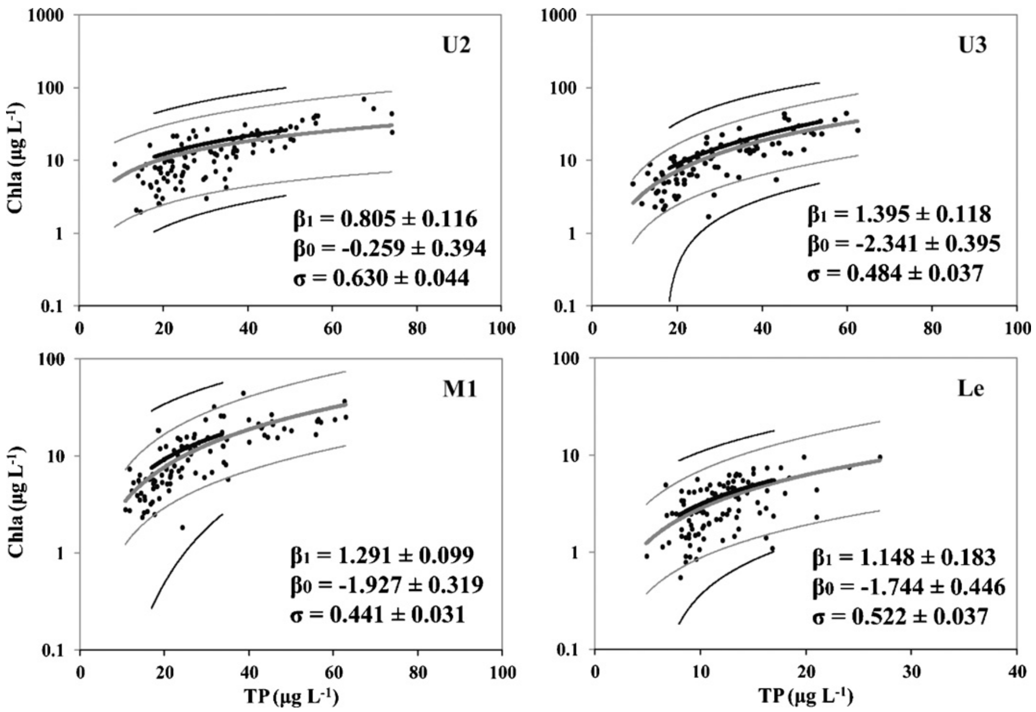

Fig. 5. The *TP*-chlorophyll *a* relationships in different segments of the Bay of Quinte. The grey and black lines correspond to the linear regression models with predictors the measured *TP* concentrations and the daily outputs from the modified (PM_{2012}) *TP* model, respectively. The black circles represent the observation data. Thick and thin lines represent the mean predictions and 95% credible intervals, respectively. The posteriors of the stochastic nodes (β_0 (intercept), β_1 (slope), and σ (standard error of the estimate)) of the empirical chlorophyll a equations were reported in each panel (mean ± standard deviation).

question, segment-specific chlorophyll *a* vs *TP* linear regression models were developed based on data from individual samplings during the 2002–2009 period, which were then combined with the corresponding outputs from the *TP* model to project daily chl*a* concentrations (Figs. 5 and 6). The integration of the two models was founded upon a Bayesian configuration to describe probabilistically the relationships among the water quality variables; an essential characteristic if predictions are to be used to guide decision-making. In particular, the empirical chlorophyll *a* equations were formulated as errors-in-variables models of the form: $\ln(\text{chl}a) = \beta_1 \cdot \ln(TP^*) + \beta_0$, where TP^* denotes the *true* value of the regressor. The latter variable was assumed to be a draw from a normal distribution, in which the mean value and standard deviation were the segment-specific daily predictions of the *TP* mass-balance model and the corresponding structural error, respectively (Berkson, 1950; Carroll et al., 2006). With this statistical expression, we allowed the propagation of the process error of the *TP* model through the predictive chlorophyll *a* distributions. We also used Cholesky decomposition of the parameter variance–covariance matrix to accommodate the dependence structure implicit in the joint posterior distribution of the regression parameters β_0 and β_1 (Golub and Van Loan, 1983). Relative to the conventional regression models, our Bayesian network provided a broader predictive range of chlorophyll *a* concentrations for a given ambient *TP* level (Fig. 5). Yet, the domain captured by our integrated modelling construct is much narrower than the observed range due to the inability of the *TP* mass-balance model to reproduce several of the observed peaks during the growing season. Thus, whilst the projected chlorophyll *a* distributions do not differ significantly from those derived by the conventional regression models (Fig. 6), we caution that these predictive statements are based on a weak representation of the chl*a*–*TP* causal association and are partly confounded with the *P* model structural error and input uncertainty.

4. Discussion

After four decades of experience with model-based water quality management, the modelling literature emphatically advocates the establishment of a systematic protocol for model development (Arhonditsis, 2009). Yet, despite the convincing presentation in several classic modelling textbooks of “rational model development” (e.g., Chapra, 1997; Jørgensen and Bendoricchio, 2001), Arhonditsis and Brett (2004) reported disturbing methodological inconsistencies in contemporary modelling practices. The large majority of the published studies in the field of aquatic biogeochemical modelling over the last decade did not rigorously quantify model sensitivity to the input vectors, while aquatic ecosystem modellers do not always examine the ability of their models to support predictions in the extrapolation domain or even the goodness-of-fit to the observed data during model calibration; see Fig. 2 in Arhonditsis and Brett (2004). In the context of environmental management, our thesis is that the typical steps involved in the development of a mathematical model, such as the predictive validation, structural confirmation, sensitivity and/or uncertainty analysis, are analogous to the way a chemical analyst strives to attain clean laboratory conditions, precise standardization curves, and meticulous application of an analytical protocol. This premise underlies our evaluation of the Minns and Moore (2004) model as a decision making tool in the Bay of Quinte.

How reliable is the existing TP model to guide future management decisions? Modelling textbooks emphasize that the calibration of a model (or the “model training” phase) does not provide any information in regard to its predictive power, but merely examines the ability of a specific model structure to match a single dataset (Chapra, 1997). It is recommended that the calibration should always be followed by the predictive evaluation; a procedure whereby the modeller tests the model against an independent set of data, which ideally should be significantly different from the one used during the calibration phase (Reckhow and Chapra, 1999).

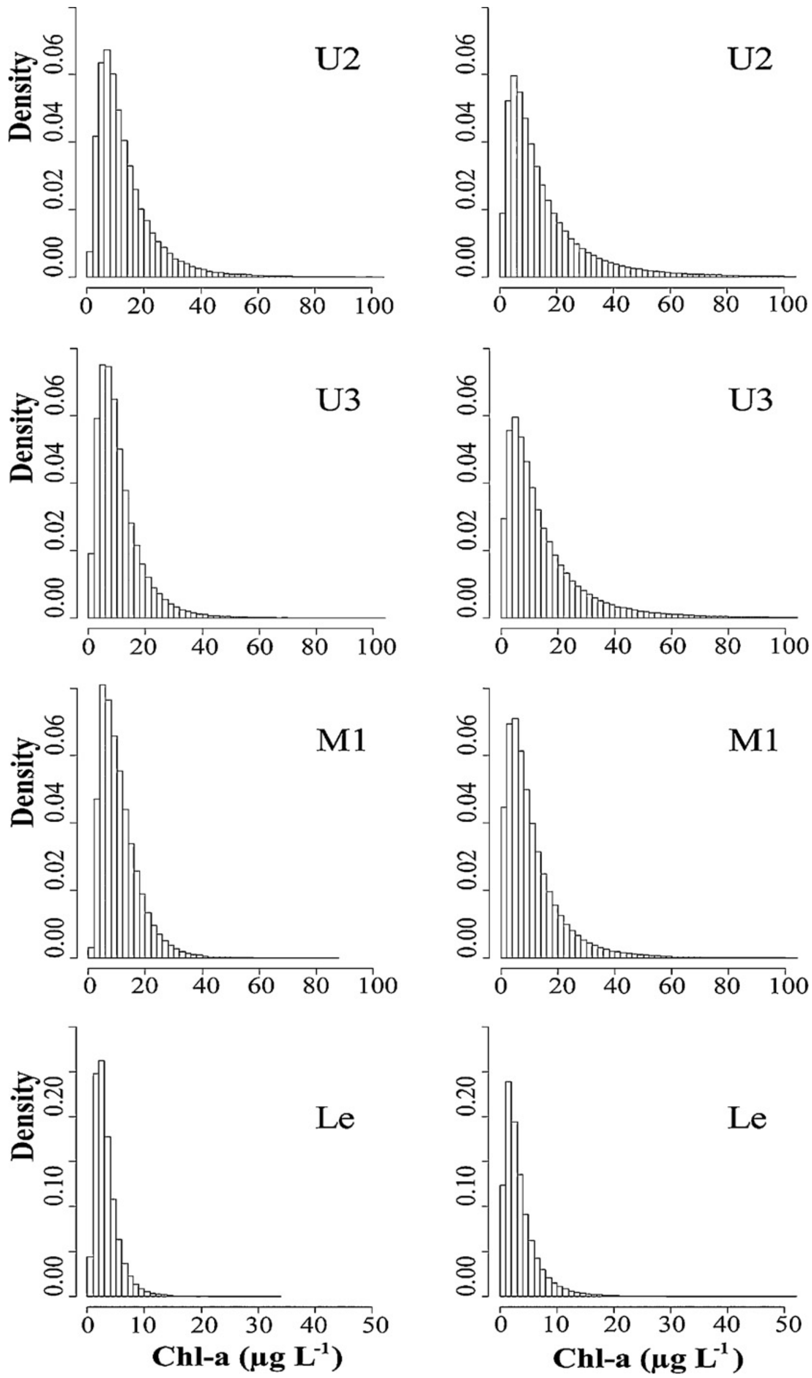

Fig. 6. Predicted from conventional regression (left panels) versus predicted from network of empirical models (right panels) chlorophyll *a* distributions at the U_2, U_3, M_1 and L_e segments during the growing season of 2008.

This phase is also referred to as model "validation", although this term may not be appropriate for models that deal with open systems and numerous sources of uncertainty (Oreskes et al., 1994). In this regard, the Bay of Quinte model is a classic case of a (nearly) excellent fit to the calibration dataset that fails to reproduce the observed patterns in the validation domain. In particular, both the r^2 and *RE* values of the original calibration lie within the top quartile of the performance typically reported for this class of models when simulating nutrient dynamics (see Fig. 3 in Arhonditsis and Brett, 2004). Yet, the original model not only demonstrated limited capacity to mimic the year-to-year variability of the 2002–2009 period, but was also characterized by a systematic underestimation bias of the observed *TP* concentrations (e.g., see the negative average error values in Table 2). The general tendency for the model to systematically understate the dynamic range of the real system is particularly evident when focus is shifted to monthly predictions.

Our analysis showed that the underestimation bias most likely stemmed from the numerically unbalanced sediment characterization of the original calibration exercise that assigned an excessively high permanent phosphorus loss from the active sediment layer. This parameterization was an effective calibration strategy to simultaneously match the high *TP* levels in the early/mid-1970s and the immediate response of the system to the substantial reduction of point-source loading, but it unrealistically moderated the replenishment of the water column in recent years by the sole feedback mechanism included in the model, i.e., the sediment

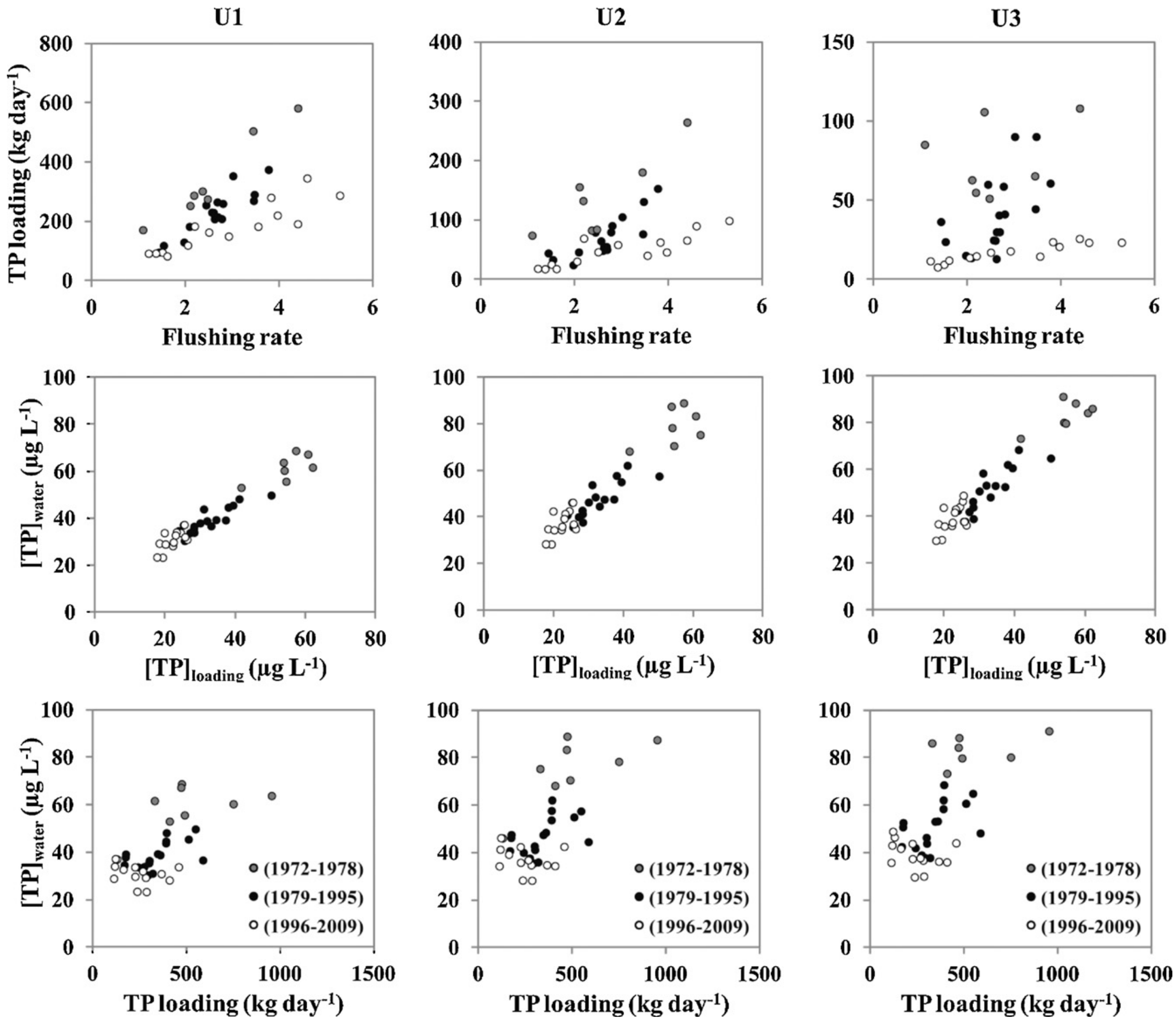

Fig. 7. Relationships between seasonal *TP* loadings and flushing rates (top row); seasonal ambient *TP* concentration and *TP* flow-weighted input concentration (middle row); seasonal *TP* concentration and *TP* loading (bottom row) in the Upper Bay. The grey circles represent the pre-phosphorus control period (1972–1978). The solid back circles represent the after P-control and the pre-dreissenids period (1979–1995), while the hollow circles represent the post-dreissenids period (1996–2009).

reflux. It also assumes a uniform sediment P content across the Bay; not only is this unlikely, given the significant among site differences in flushing rates and productivity, but recent field data shows significant variance in sediment composition and nutrient content among sites (Watson and Guo, unpublished data). Explicit consideration of temperature-dependent sediment release rates and parameterization with the PM_{2012} version addressed some of these issues, and improved both seasonal and monthly predictive statements of the model. While this is a promising result, there are two basic reasons that call into question the general applicability of this particular model construct: (i) the model still fails to consistently reproduce the end-of-summer *TP* accumulation in the middle and lower segments of the Bay of Quinte; and (ii) a careful inspection of the right panels in Fig. 3 suggests that the improvement in the model fit was achieved through a global upward shift of the baseline of the ambient *TP* levels, driven by the intensification of the internal nutrient loading, whereas the amplitude of the simulated *TP* oscillations remained smaller than the observed intra-annual variability in the system.

Modellers are commonly confronted with what is called an “inverse” problem: there is sufficient information on the levels and the variability of the state (or output) variables, but little is known about the values of the model input parameters (Oreskes et al., 1994). This problem is typically resolved by adjusting the parameters to find the best agreement between model outputs and observed data, and the derived model parameterization is subsequently used to draw inference about the postulated ecosystem functioning (Arhonditsis and Brett, 2005). In this regard, both our *TP* budget and sensitivity analysis indicated that the Trent River loading overwhelmingly dominates the dynamics of the upper Bay until the main stem of the middle area. However, the actual nature of the relationship between exogenous loading forcing and system response is less straightforward and invites further investigation. In particular, we examined the temporal alterations of the relationships among the water flushing rates, *TP* input concentrations, *TP* loadings, and the seasonal *TP* levels in the upper Bay of Quinte, in response to the reduction of exogenous nutrient loading and subsequently the invasion of dreissenids (Fig. 7). First, we highlight the general pattern of a positive causal association between *TP* loadings and flushing rates in the three upper segments over the entire time span of our analysis. Yet, the reduced slope in most recent years probably reflects a gradual disconnect between the hydraulic loading and the actual mass of phosphorus entering the system. Flow-weighted *TP* input concentration and *TP* loading are both characterized by a positive relationship with the *TP* levels in the upper Bay. Notably, the relationship between ambient and exogenous concentrations maintains its monotonic character throughout the study period, whereas the relationship with the external loading was distinctly weaker (or even demonstrated a negative trend) during the 1996–2009 period. The latter pattern highlights the confounding, or even conflicting, effects of the two components of loading (flow and concentration) on attributes of concern. Much like other advective systems (Borsuk et al., 2004; Boynton and Kemp, 2000; Dettmann, 2001), the reconciliation of the direct control of exogenous loading on the ambient *TP* concentrations vis-à-vis the indirect effects mediated by the flow-controlled downstream transport is critical for properly predicting the upper Bay of Quinte response to watershed management actions (see also following discussion).

According to the model predictions, the reflux of phosphorus from the sediments represents a pathway that becomes increasingly important as we move away from the mouth of Trent River. Because this result refers to the only feedback loop considered by the model, we believe that it collectively reflects the importance of all the mechanisms of internal nutrient loading/recycling and may not be associated with the role of the sediments per se. Given that the fairly consistent end-of-summer *TP* accumulation in the system cannot be fully explained by the known loading contributions of the exogenous sources during the summer period, it seems reasonable to assume that the nutrient regeneration mechanisms could partly modulate the ambient *TP* levels. The likelihood of substantial phosphorus fluxes from the sediments, especially in the post-dreissenid period, has been discussed in the literature (Minns and Moore, 2004; Minns et al., 2004), although this hypothesis has yet to be rigorously examined in the field. Our modelling analysis builds upon the assertions of earlier work in the Bay of Quinte area, suggesting that (i) the phosphorus fluxes emanating from internal sources are quantitatively comparable to the exogenous loading in the upper Bay, and (ii) may also predominantly drive the *TP* dynamics in enclosed embayments, such as the Hay Bay (M_2), and Picton Bay (M_3). The former result in conjunction with the aforementioned capacity of the hydraulic loading from Trent River to regulate the residence time in the upper Bay highlights a complex interplay between internal recycling and local hydrodynamic regime (e.g., low flushing conditions) that may be responsible for the end-of-summer high *TP* concentrations in Big Bay or Muscote Bay (U_2). The latter finding though postulates a greater reliance of the ecosystem functioning upon nutrient recycling mechanisms with the increasing distance from Trent River, which is conceptually on par with the spatiotemporal changes of the lower food web structure reported by Munawar et al. (2011). Yet, we caution that the exogenous loading estimates in the two embayments are quite uncertain, as they are entirely based on Wilton Creek equivalents, and thus the derived *TP* budgets may unrealistically overstate the importance of nutrient regeneration mechanisms in these segments.

Are we targeting the appropriate water quality criteria? Hitherto, the evaluation of the phosphorus mass-balance model suggests that the reparameterized PM_{2012} version with finer segmentation and temperature-dependent sediment reflux rates could provide an adequate management tool, if we are striving for predictive statements at the seasonal scale. Such temporal resolution should seemingly meet the demands of the policy-making process, given that all the historical restoration objectives in the Bay were expressed as "average values during the growing season" (Bay of Quinte RAP, 1993). The question arising though is to what extent a seasonally averaged value, derived from samples collected on a biweekly basis from one or two offshore sites, can impartially represent the real water quality issues in the Bay of Quinte area? Recent work reports *TP* concentrations in locations of high public exposure (e.g., beaches) as well as in uncharted river mouths, shorelines/embayments, and sites near municipal inputs that far exceed levels typically encountered in the long-term monitoring sites, indicative of a dramatic difference in the water quality between inshore and offshore areas (Watson et al., 2011). The latter pattern may also have broader implications about the appropriate spatiotemporal resolution for studying and subsequently drawing inference on the evolution of the ecosystem state. For example, there were instances in which samples from downstream of the local wastewater treatment plants or the midstream offshore waters in Trenton and Picton Harbour ranged up to 800 $\mu g\ TP\ L^{-1}$. Interestingly, all samples with *TP* greater than 80 $\mu g\ L^{-1}$ were collected from the upper surface and were also associated with high chlorophyll *a* and particulate matter concentrations. In this regard, Watson et al. (2011) asserted that nutrients bound in this buoyant particulate material can be rapidly transported over large areas by wind and wave action, and thus the *TP* concentrations may be subject to considerable variations in both space and time.

The apparent disconnect between water quality at long term monitoring sites and shoreline events is particularly troublesome and resonates with other skeptical views about the recovery pace of the system (Bowen and Johannsson, 2011; Minns et al., 2011; Munawar et al., 2011). Watson et al. (2011) provided evidence that these pronounced spatial heterogeneity patterns partly stem from the local influence of wastewater (or urban storm sewer) discharges. Given that the seasonal-average (May–October) point source loading in the Upper Bay has been lower than 15 kg day^{-1} during the model validation period (2002–2009), this finding implies that phosphorus loading from bypassing events, typically not quantified in the Bay of Quinte water quality reports, could be appreciable and may add significantly to the phosphorus loading (Kinstler and Morley, 2011). Another plausible explanation may be associated with the credibility of the current non-point loading estimates that are based on questionable proxies (e.g., *WCE*s) to capture the contribution of un-gauged basins and/or are overwhelmingly biased towards low flow samples (see also companion paper by Kim et al., 2013). In light of recent studies establishing the significant role of event flows in the annual load determination (Macrae et al., 2007), we believe that this is a major knowledge gap making compelling the design of a nutrient concentration sampling programme that targets event flows (Pollak et al., 2013). A characteristic example is a recent study by Labencki and coworkers in the Hamilton Harbour watershed, which offered significant insights concerning the contingency of the annual load estimation on the temporal sampling protocol. Namely, there was evidence that the majority of the total phosphorus export occurs under event flow conditions, and therefore the conventional monitoring programmes that primarily focus on baseline conditions may significantly underestimate phosphorus loading at both urban and agriculture catchments (Macrae et al., 2007; Wellen et al., 2012).

To recap, existing empirical evidence makes it abundantly clear that the currently targeted seasonal average *TP* concentrations of 30 $\mu g\ L^{-1}$ is neither representative of the water quality conditions in areas of high public exposure (e.g., beaches) nor does it reflect the actual temporal variability in the system. Paradoxically, while both pillars of model-based environmental management seem to be in place, i.e., an attainable water quality target and a mathematical model that can support predictions for the spatiotemporal resolution level required, the process itself seems to be disengaged from the actual water quality problems of the Bay of Quinte. In this regard, we believe that the new *TP* delisting objective and model predictive capacity should be framed upon: (i) a finer temporal scale, such as monthly or even daily snapshots from samplings that cover the entire growing season; (ii) the need to accommodate the considerable spatial variability in the system; (iii) the importance of identifying a numerical value that is both scientifically sound and achievable; and (iv) the pragmatic view that the criteria setting process should explicitly accommodate the natural variability of the system or the substantial uncertainty characterizing the existing exogenous loading estimates by permitting a realistic frequency of goal violations. For example, an exceedance frequency of 20% or less of the samples collected during the focal period (i.e., May–September) should still be considered system compliance. In a follow-up study, we present an analysis of nutrient loading scenarios that aims to identify which of the above conditions can be addressed unequivocally (Kim et al., 2013).

What is the most suitable mechanistic augmentation of the existing model? From a management standpoint, our study highlights the proper representation of the causal association among exogenous loading, internal recycling, and end-of-summer ambient concentrations as one of the key challenges of *TP* modelling in the Bay of Quinte. In this regard, aside from the aforementioned need to improve the existing loading estimates, we believe that the explicit mathematical depiction of factors that mediate the internal nutrient recycling are likely to offer mechanistic insights into the ambient *TP* accretion during the summer period. Quite recently, Ramin et al. (2012) reinforced the (oftentimes neglected) notion that nutrient recycling processes can significantly influence the predictions of any aquatic biogeochemical modelling exercise, as the recycled material and the associated nutrient fluxes can be significant drivers in low as well as in high productivity ecosystems. Namely, the non-predatory algal or macrophyte death along with the subsequent release of stored inorganic phosphorus, the decomposition and mineralization activity of individual groups (bacteria, heterotrophic nanoflagellates) against a variety of substrates (e.g., macrophytes, living and/or dead algal cells) with different nutritional/biochemical content, the dreissenid production of pseudofeces, the animal-mediated metabolic recycling and translocation can potentially be significant pathways of the phosphorus cycle, even in lakes that receive high external loadings (Bierman et al., 2005; Kamarainen et al., 2009; Ramin et al., 2012; Vanni, 2002). In this context, the incorporation of submodels that focus on the seasonal variability of macrophyte growth (Leisti et al., 2006), the functional role of dreissenid mussels (Minns et al., 2011), the fate and transport of phosphorus in the sediments are likely to improve our contemporary understanding of the system and subsequently the model capacity to reproduce the substantial intra-annual *TP* variability in the Bay of Quinte (see also companion paper by Kim et al., 2013).

Similar to Minns et al.'s (2004) assertions, another important facet of the present model construct involves the assumptions made about the backflows from Lake Ontario, which appear to be particularly influential on the phosphorus budgets in the middle and lower segments of the Bay (see Table 3). During the calibration of the PM_{2012} model, we found that the dilution effects of the Lake Ontario water masses could profoundly modulate the seasonality patterns in the middle segments (M_2 and M_3), the hypolimnetic *TP* accrual (L_h) as well as the vertical entrainment of phosphorus-rich water masses in the epilimnion of the lower segment (L_e). The ongoing *ELCOM* application in the Bay of Quinte will certainty shed light on the spatial and temporal variability of the hydrodynamic conditions (Oveisy et al., 2012), although we caution that the integration of the simulated circulation profiles with the present model segmentation may not be a straightforward exercise. Earlier work by Shanahan and Harleman (1984) contended that in order to minimize the implicit dispersion introduced by the discrete spatial structure, exchanges in multiple-box models should only be derived via calibration and cannot be directly determined by the hydrodynamics of the prototype system. While the prospect of an additional calibration, alongside the parameterization of the *TP* model, certainly entails an inflation of the model uncertainty, we believe that the presence of a more sophisticated hydrodynamic model will be beneficial for two basic reasons: (i) it will allow verifying the plausibility of the residence times postulated by the present models; and (ii) it will offer the capacity to more sensibly examine the local response of small embayments (e.g., Muscote Bay, Picton Bay, Hay Bay) to episodic meteorological perturbations (extreme precipitation events) or wastewater discharges that may intermittently disconnect the water quality between inshore and offshore sites in the Bay of Quinte.

On a final note, the most common misinterpretation of Occam's razor is that "the simplest model is most likely a better one", although what the principle actually suggests is a shift towards simpler theories until simplicity can be gradually traded for increased predictive capacity (Jaynes, 1994). In this context, our recommendations invoke extra ecological complexity and finer spatial segmentation to assess the adequacy of the Minns and Moore (2004) model, but do not suggest that the development of an overly complex model by itself is the "panacea" for achieving robust a management tool! In fact, the increase of the ecological (expressed as the number of state variables) or the spatial (from zero- to three-dimensional approaches) model complexity does not necessarily improve model performance (e.g., see Table 2 in Arhonditsis and Brett, 2004). Rather, decisions on the complexity of a model should be driven by the system under study and the questions asked. In essence, simple ecological models can offer sensible first-order approximations to establish a realistic representation of the causal connections among exogenous nutrient loading, ambient nutrient conditions, and lower food web dynamics, i.e., the factors primarily associated with the manifestation of eutrophication problems. Added complexity (more biotic compartments or a three-dimensional hydrodynamic approach) should only be considered when there is evidence that our explanatory power will increase, the available information from the system can reasonably constrain the model, and the resulting modelling construct will not impede our ability to rigorously quantify the underlying uncertainty (Arhonditsis, 2009; Arhonditsis et al., 2007, 2008a, 2008b). If any of the three (not mutually exclusive) conditions is not met, then the adoption of a complex model may considerably compromise our ability to sensibly guide the management of our water resources.

Acknowledgments

This project was undertaken with the financial support of the Lower Trent Region Conservation Authority provided through the Bay of Quinte Remedial Action Plan Restoration Council. 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. We gratefully acknowledge the insightful comments from Ken Minns and Scott Millard on an earlier version of the manuscript.

References

Arhonditsis, G.B., 2009. Useless arithmetic? Lessons learnt from aquatic biogeochemical modelling. In: Hanrahan, G. (Ed.), Modelling of Pollutants in Complex Environmental Systems. ILM Publications, pp. 3–26.

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

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

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

Arhonditsis, G.B., Perhar, G., Zhang, W., Massos, E., Shi, M., Das, A., 2008b. Addressing equifinality and uncertainty in eutrophication models. Water Resources Research 44, 19.

Arhonditsis, G.B., Qian, S.S., Stow, C.A., Lamon, E.C., Reckhow, K.H., 2007. Eutrophication risk assessment using Bayesian calibration of process-based models: application to a mesotrophic lake. Ecological Modelling 208, 215–229.
Bay of Quinte RAP, 1993. Stage 2 Report: Time to Act. The Bay of Quinte Remedial Action Plan, p. 253.
Berkson, J., 1950. Are there two regressions? Journal of the American Statistical Association 45, 164–180.
Bierman, V.J., Kaur, J., DePinto, J.V., Feist, T.J., Dilks, D.W., 2005. Modeling the role of zebra mussels in the proliferation of blue-green algae in Saginaw Bay, Lake Huron. Journal of Great Lakes Research 31, 32–55.
Borsuk, M., Stow, C., Reckhow, K., 2004. Confounding effect of flow on estuarine response to nitrogen loading. Journal of Environment Engineering 130, 605–614.
Borsuk, M.E., Stow, C.A., Reckhow, K.H., 2002. Predicting the frequency of water quality standard violations: a probabilistic approach for TMDL development. Environmental Science and Technology 36, 2109–2115.
Bowen, K.L., Johannsson, O.E., 2011. Changes in zooplankton biomass in the Bay of Quinte with the arrival of the mussels, *Dreissena polymorpha* and *D. rostiformis bugensis*, and the predatory cladoceran, *Cercopagis pengoi*: 1975 to 2008. Aquatic Ecosystem Health & Management 14, 44–55.
Boynton, W.R., Kemp, W.M., 2000. Influence of river flow and nutrient loads on selected ecosystem processes: a synthesis of Chesapeake Bay data. In: Hobbie, J.E. (Ed.), Estuarine Science: A Synthetic Approach to Research and Practice. Island Press, Washington, DC, pp. 269–298.
Carroll, R.J., Ruppert, D., Stefanski, L.A., Crainiceanu, C.M., 2006. Measurement Error in Nonlinear Models: A Modern Perspective. Chapman & Hall/CRC, Oxford, UK, 455 pp.
Chapra, S.C., 1997. Surface Water-Quality Modeling. McGraw-Hill, New York, 844 pp.
Christie, W.J., Becker, M., Cowden, J.W., Vallentyne, J.R., 1986. Special contribution: managing the Great Lakes basin as a home. Journal of Great Lakes Research 12, 2–17.
Dettmann, E., 2001. Effect of water residence time on annual export and denitrification of nitrogen in estuaries: a model analysis. Estuaries and Coasts 24, 481–490.
Diamond, M.L., Mackay, D., Poulton, D.J., Stride, F.A., 1996. Assessing chemical behavior and developing remedial actions using a mass balance model of chemical fate in the Bay of Quinte. Water Research 30, 405–421.
Golub, G.H., Van Loan, C.F., 1983. Matrix Computations. Johns Hopkins University Press, Baltimore, MD, 699 pp.
Jaynes, E.T., 1994. Probability Theory: The Logic of Science. Cambridge University Press, New York.
Jørgensen, S.E., Bendoricchio, G., 2001. Fundamentals of Ecological Modelling. Elsevier Science Ltd., Amsterdam, 544 pp.
Kamarainen, A., Penczykowski, R., Van de Bogert, M., Hanson, P., Carpenter, S., 2009. Phosphorus sources and demand during summer in a eutrophic lake. Aquatic Sciences: Research Across Boundaries 71, 214–227.
Kim, D.-K., Zhang, W., Rao, Y., Watson, S., Mugalingam, S., Labencki, T., Dittrich, M., Morley, A., Arhonditsis, G.B., 2013. Improving the representation of internal nutrient recycling with phosphorus mass balance models: a case study in the Bay of Quinte, Ontario. Canada. Ecological Modelling 256, 53–68.
Kinstler, P., Morley, A., 2011. Point Source Phosphorus Loadings 1965 to 2009, Bay of Quinte Remedial Action Plan: Monitoring Report #20, Kingston, Ontario, Canada, pp. 15–17.
Koops, M.A., Irwin, B.J., MacNeil, J.E., Millard, E.S., Mills, E.L., 2006. Comparative Modelling of the Ecosystem Impacts of Exotic Invertebrates and Productivity Changes on Fisheries in the Bay of Quinte and Oneida Lake. Completion Report, Fisheries Research Program, Great Lakes Fishery Commission.
Leisti, K.E., Millard, E.S., Minns, C.K., 2006. Assessment of submergent macrophytes in the Bay of Quinte, Lake Ontario, August 2004, Including Historical Context. Department of Fisheris and Oceans Canada, p. 81.
Macrae, M.L., English, M.C., Schiff, S.L., Stone, M., 2007. Capturing temporal variability for estimates of annual hydrochemical export from a first-order agricultural catchment in southern Ontario, Canada. Hydrological Processes 21, 1651–1663.
Manache, G., Melching, C.S., 2004. Sensitivity analysis of a water-quality model using Latin hypercube sampling. Journal of Water Resources Planning and Management (ASCE) 130, 232.
Minns, C.K., Hurley, D.A., Nicholls, K.H., 1986. Project Quinte: point-source phosphorus control and ecosystem response in the Bay of Quinte, Lake Ontario. Department of Fisheries and Oceans, Ottawa.
Minns, C.K., Moore, J.E., 2004. Modelling Phosphorus Management in the Bay of Quinte, Lake Ontario in the Past, 1972 to 2001, and in the Future, Canadian Manuscript Report of Fisheries and Aquatic Sciences No. 2695, Burlington, Ontario, p. 42.
Minns, C.K., Moore, J.E., Seifried, K.E., 2004. Nutrient loads and budgets in the Bay of Quinte, Lake Ontario, 1972 to 2001. Canadian Manuscript Report of Fisheries and Aquatic Sciences 2694, Burlington, Ontario, p. 49.
Minns, C.K., Munawar, M., Koops, M.A., Millard, E.S., 2011. Long-term ecosystem studies in the Bay of Quinte, Lake Ontario, 1972–2008: a prospectus. Aquatic Ecosystem Health & Management 14, 3–8.
Munawar, M., Fitzpatrick, M., Niblock, H., Lorimer, J., 2011. The relative importance of autotrophic and heterotrophic microbial communities in the planktonic food web of the Bay of Quinte, Lake Ontario 2000–2007. Aquatic Ecosystem Health & Management 14, 21–32.
Office of Water, 1997. Guidelines for Preparation of the Comprehensive State Water Quality Assessments. Environmental Protection Agency, Washington DC.
Oreskes, N., Shrader-Frechette, K., Belitz, K., 1994. Verification, validation, and confirmation of numerical models in the earth sciences. Science 263, 641–646.
Oveisy, A., Boegman, L., Imberger, J., 2012. Three-dimensional simulation of lake and ice dynamics during winter. Limnology and Oceanography 57, 43–57.
Pollak, A.W., Peirce, J.J., Alvarez-Vazquez, L.J., Vazquez-Mendez, M.E., 2013. Methodology for identifying optimal locations of water quality sensors in river systems. Environmental Modeling & Assessment 18, 95–103.
Ramin, M., Labencki, T., Boyd, D., Trolle, D., Arhonditsis, G., 2012. A Bayesian synthesis of predictions from different models for setting water quality criteria. Ecological Modelling 242, 127–145.
Reckhow, K.H., Arhonditsis, G.B., Kenney, M.A., Hauser, L., Tribo, J., Wu, C., Elcock, K.J., Steinberg, L.J., Stow, C.A., McBride, S.J., 2005. A predictive approach to nutrient criteria. Environmental Science and Technology 39, 2913–2919.
Reckhow, K.H., Chapra, S.C., 1999. Modeling excessive nutrient loading in the environment. Environmental Pollution 100, 197–207.
Shanahan, P., Harleman, D., 1984. Transport in lake water quality modeling. Journal of Environment Engineering 110, 42–57.
Stow, C.A., Roessler, C., Borsuk, M.E., Bowen, J.D., Reckhow, K.H., 2003. Comparison of estuarine water quality models for total maximum daily load development in Neuse River estuary. Journal of Water Resources Planning and Management (ASCE) 129, 307–314.
Taraborelli, A.C., Fox, M.G., Johnson, T.B., Schaner, T., 2010. Round Goby (*Neogobius melanostomus*) population structure, biomass, prey consumption and mortality from predation in the Bay of Quinte, Lake Ontario. Journal of Great Lakes Research 36, 625–632.
Vanni, M.J., 2002. Nutrient cycling by animals in freshwater ecosystems. Annual Review of Ecology and Systematics 33, 341–370.
Watson, S.B., Borisko, J., Lalor, J., 2011. Bay of Quinte Harmful Algal Bloom Programme. Phase I – 2009, Bay of Quinte Remedial Action Plan: Monitoring Report #20, Kingston, Ontario, Canada, pp. 27–50.
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 Resources Research 48, W10505.
Zhang, W., Arhonditsis, G.B., 2008. Predicting the frequency of water quality standard violations using Bayesian calibration of eutrophication models. Journal of Great Lakes Research 34, 698–720.

Glossary

Areas of Concern (AOCs): refers to geographic areas within the Great Lake system that are characterized by severe environmental degradation and fail to support one or more of their beneficial uses.

Beneficial Use Impairments (BUIs): a change in the physical, chemical, or biological integrity of a water body sufficient to impair any of the fourteen (14) beneficial uses identified by the International Joint Commission (*IJC*) for listing and delisting Areas of Concern in the Great Lakes Basin Ecosystem. The beneficial use impairments are as follows: restrictions on fish and wildlife consumption; tainting of fish and wildlife flavour; degradation of fish and wildlife populations; fish tumours or other deformities; bird or animal deformities or reproduction problems; degradation of benthos; restrictions on dredging activities; eutrophication or undesirable algae; drinking water restrictions, or taste and odour problems; beach closings; degradation of aesthetics; added costs to agriculture or industry; degradation of phytoplankton and zooplankton; loss of fish and wildlife habitat.

Ecosystem management approach: a paradigm founded upon the premise that the changes in the ecosystem integrity have occurred in the Great Lakes basin as the result of the cumulative effects of many local stressors and are not attributable to any specific factor. The ecosystem management approach aims to manage all local manmade stressors in the belief that this will result in the restoration of the impaired system.

International Joint Commission (IJC): an independent binational organization established by the United States and Canada under the Boundary Waters Treaty of 1909 in order to manage water resources wisely and to protect them for the benefit of today's citizens and future generations.

Listing/delisting criterion: refers to water quality standards that serve as measurable surrogates for the prevailing conditions in a particular waterbody. The criteria aim to provide easily measurable and good predictors of the ecosystem restoration and maintenance efforts. They have been used as indicators of Beneficial Use Impairments (*BUIs*) for Great Lakes Areas of Concern (*AOCs*).

Percentile-based water quality standards: a probabilistic assessment of the prevailing conditions that allows for a pre-specified frequency level of violations of targeted water quality goals.

Remedial Action Plans (RAP): are usually formulated through a wide variety of government, private sector, and community participants to provide the framework for actions aimed at restoring Areas of Concern. All *RAP*s must proceed through three stages. Stage One determines the severity and underlying causes of environmental degradation that make the location an Area of Concern. Stage Two identifies goals and recommends actions that will lead to the restoration and protection of ecosystem health. Stage Three implements recommended actions and measures progress of restoration and protection efforts in the Area of Concern to ensure the local goals have been met.