International Environmental Modelling and Software Society (iEMSs)
2012 International Congress on Environmental Modelling and Software
Managing Resources of a Limited Planet, Sixth Biennial Meeting, Leipzig, Germany
R. Seppelt, A.A. Voinov, S. Lange, D. Bankamp (Eds.)
http://www.iemss.org/society/index.php/iemss-2012-proceedings

Integration of Bayesian inference techniques with mathematical modelling

George B. Arhonditsis
Ecological Modelling Laboratory, Department of Physical & Environmental Sciences, University of Toronto, Toronto, Ontario, M1C 1A4, Canada
e-mail: georgea@utsc.utoronto.ca

Abstract: Skeptical views of the scientific value of modelling argue that there is no true model of an ecological system, but rather several adequate descriptions of different conceptual basis and structure. My study addresses this question using a complex ecosystem model, developed to guide the water quality criteria setting process in the Hamilton Harbour (Ontario, Canada), along with a simpler plankton model that considers the interplay among phosphate, detritus, and generic phytoplankton and zooplankton state variables. Predictions from the two models are combined using the respective standard error estimates as weights in a weighted model average. The two eutrophication models are used in conjunction with the SPAtially Referenced Regressions On Watershed attributes (SPARROW) watershed model. The Bayesian nature of my work is used: (i) to alleviate problems of spatiotemporal resolution mismatch between watershed and receiving waterbody models; and (ii) to overcome the conceptual or scale misalignment between processes of interest and supporting information. The lessons learned from this study will contribute towards the development of integrated modelling frameworks.

***Keywords**:* Process-based modelling, Eutrophication, Bayesian inference, Water quality criteria, Decision making.

1. Introduction

In the context of water quality assessment, the application of process-based models typically has a deterministic character, whereby single-value predictions at each point in time and space are derived from uniquely determined model inputs. Most of the existing calibration efforts aim at reproducing the average ecological dynamics, but fail to capture the entire range of natural conditions experienced. The credibility of these practices and their adequacy in addressing environmental management problems has recently been questioned for two main reasons [Arhonditsis et al. 2007]. First, regardless of its complexity and supporting information, the application of any modeling construct involves substantial uncertainty contributed by model structure, parameters, and other associated inputs (e.g., boundary or initial conditions). Second, models parameterized to depict the average ecosystem behavior are inadequate in addressing the type of percentile-based standards needed to accommodate the natural spatiotemporal variability and may bias (underestimate) the predictions of the frequency of standard violations under various management options [Borsuk et al. 2002].

For better model-based decision analysis that can effectively support the development of environmental standards and the policy making process, the uncertainty in model predictions as well as the full range of the expected system responses must be rigorously quantified and reported in a straightforward way. Model uncertainty analysis essentially aims to make inference about the joint probability distribution of model inputs, reflecting the amount of knowledge available for model parameters, initial conditions, forcing functions, and model structure. In this regard, Bayes' Theorem provides a convenient means to combine existing information (prior) with current observations (likelihood) for projecting future ecosystem response (posterior). Hence, the Bayesian techniques are more informative than the conventional model calibration practices, and can be used to refine our knowledge of model input parameters while obtaining predictions along with uncertainty bounds for output variables [Arhonditsis et al. 2007].

Despite the compelling arguments for considering Bayesian inference techniques as an integral part of the model development process, their high computational demands along with the lack of analytical expressions for the posterior distributions was until recently a major impediment for their broader application. Nonetheless, the advent of fast computing has allowed the development of several methods for performing Bayesian inference and the most commonly used technique is called Markov chain Monte Carlo (MCMC); a general methodology that provides a solution to the difficult problem of sampling from high dimensional distributions for the purpose of numerical integration. In this paper, I will discuss several promising prospects of the application of Bayesian inference techniques, such as the averaging of predictions from different models and the integration of watershed with receiving waterbody models, which can be used from stakeholders and policy makers to guide the use of millions of dollars of restoration and to dictate the Best Management Practices.

2. Case study

Hamilton Harbour, a large embayment located at the western end of Lake Ontario, has a long history of eutrophication problems primarily manifested as excessive algal blooms, low water transparency, predominance of toxic cyanobacteria, and low hypolimnetic oxygen concentrations during the late summer [Gudimov et al. 2011]. Since the mid 80s, when the Harbour was identified as one of the 43 Areas of Concern (*AOC*) in the Great Lakes area, the Hamilton Harbour Remedial Action Plan (*RAP*) was formulated through a variety of government, private sector, and community participants to provide the framework for actions aimed at restoring the Harbour environment. 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. The drastic nutrient loading reduction has historically played a central role in the restoration efforts, although the determination of the critical levels has been a thorny issue as the population growth and increasing urbanization accentuate the pressure for expansion of the local wastewater treatment plants (*WWTP*s).

Recent modelling work suggests that the water quality goals for *TP* levels <20 $\mu g\ L^{-1}$, chlorophyll *a* concentrations between 5-10 $\mu g\ L^{-1}$, and water clarity >3 m will likely be met, if the proposed phosphorus loading reductions at the level of 142 kg day^{-1} are actually achieved [Ramin et al. 2011]. Yet, it was emphasized that the predictive capacity of any modelling exercise in the Harbour is conditional upon the credibility of the contemporary nutrient loading estimates, which are uncertain and appear to inadequately account for the contribution of non-point sources, episodic meteorological events (e.g., spring thaw, intense summer storms), and short-term variability at the local *WWTP*s. The same modelling work also pinpointed two important unknown factors that can potentially modulate the response of the system to the exogenous nutrient loading reduction and may shape the duration of the transient phase as well as the system resilience in the "post-recovery" era. First, the dynamics of phosphorus in the sediment-water column interface are still poorly understood, and thus the historical notion that the internal loading in the Harbour is minimal may be inaccurate [Gudimov et al. 2011]. Second, we lack fundamental knowledge of the regulatory factors of herbivorous zooplankton abundance and composition, even though existing evidence suggests that a thriving zooplankton community can be instrumental for achieving faster recovery rates in the Harbour. The latter prospect highlights a central conclusion drawn from my recent work that the bottom-up (i.e., nutrient loading reduction) approach historically followed in the area was sufficient to bring the system in its present state, but any further improvements should be sought in the context of a combined bottom-up and top-down control [Ramin et al. 2011].

3. Integrated modelling framework

We developed an integrated modelling framework that is founded upon *i)* a SPARROW model configuration that accommodates the interannual loading variability in the Hamilton Harbour watershed; *ii)* a Bayesian downscaling algorithm that transforms the annual nutrient loading predictions to daily estimates; and *iii)* two eutrophication models that will be used to address the following important questions regarding the future response of the system: How possible is it to meet the objective of delisting the study system as an Area of Concern, if the nutrient loading reductions proposed by the Hamilton Harbour Remedial Action Plan are actually implemented? What additional remedial actions are needed to increase the likelihood of meeting the water quality targets?

3.1 Watershed modelling

The *SPARROW* model has been extensively described elsewhere [Wellen et al., 2012], so only a basic introduction is given here. *SPARROW* is a hybrid empirical/process-based model 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. Each subwatershed is then disaggregated into reach catchments draining to a particular stream segment. Mean annual watershed export of any constituent is expressed as a function of watershed attributes. The model considers source and sink processes over annual timescales. Source processes, described with export coefficients, predict constituent mobilization; delivery factors predict how landscape attributes modulate the delivery of the mobilized constituent to streams; and attenuation coefficients predict the amount of the delivered constituent remaining in transit per length of stream or per reservoir.
In this study, Wellen et al. (2012) presented a statistical approach that introduces temporal variability to the *SPARROW* model by applying a repeated measures approach to a network of water quality monitoring stations. Rather than selecting a single year to phase out the variability in time and subsequently focusing on the spatial variability, we calibrate the model to annual loads measured repeatedly at a subset of intensively monitored sites in the studied watershed. With this statistical configuration, the *SPARROW* model is used to estimate a static baseline level of nutrient loading (μ_i) over the study period and forcing factors are being employed to explain the temporal variability around that baseline:

$$Y_{i,t} = \mu_i + W_{v,t}\gamma_v + \varepsilon_{i,t} \qquad \varepsilon_{i,t} \sim N(0,\sigma^2) \qquad \textbf{(1)}$$

where $Y_{i,t}$ refers to the natural logarithm of the measured annual load at subwatershed monitoring station *i* during year *t*, μ_i refers to a prediction of the natural logarithm of a baseline annual load at monitoring station *i* estimated by the *SPARROW* equation, $W_{v,t}$ denotes a matrix of *v, 1:V,* temporal forcing factors across years *t, 1:T,* γ_v denotes the corresponding vector of coefficients, and $\varepsilon_{i,t}$ represents an independent spatiotemporal error. All errors are assumed independent, normally distributed, and with zero mean. The temporal variability could conceivably be accommodated by anything other than watershed landscape attributes, and the focus here is on climatic factors, namely total annual precipitation and potential evapotranspiration.
The parameterization of the *SPARROW* model was based on measured loading data from the period 1988-2007 (Fig. 1; top panel). The calibration exercise offered estimates of the export coefficients and the delivery rates from the different subcatchments and thus generated testable hypotheses regarding the nutrient export "hot spots" in the watershed. We found that sites which are both large and close to the harbour have the highest delivery values per area, as the attenuation of their loads en route to the system is very low and the urban developments in the Harbour’s basin are more concentrated along the Harbour’s shore (Fig. 1; bottom panel). Further, the estimates of total phosphorus export suggested that urban land uses may export more phosphorus per area than agricultural lands. This finding is somewhat contrary to the popular notion that the rates of nutrient export from urban lands are lower than those of agricultural lands due to lower nutrient subsidies. This result may be due to the very short residence time of water in urban streams and

the limited contact runoff has with the soil matrix, which tends to trap particulate phosphorus and chemically occlude soluble phosphorus [Wellen et al., 2012]. Soil compaction due to recent construction may cause significant declines in soil infiltration capacity and a consequent increase in the generation of runoff. The higher nutrient delivery to streams in urban areas could possibly explain higher nutrient export rates despite lower nutrient subsidies.

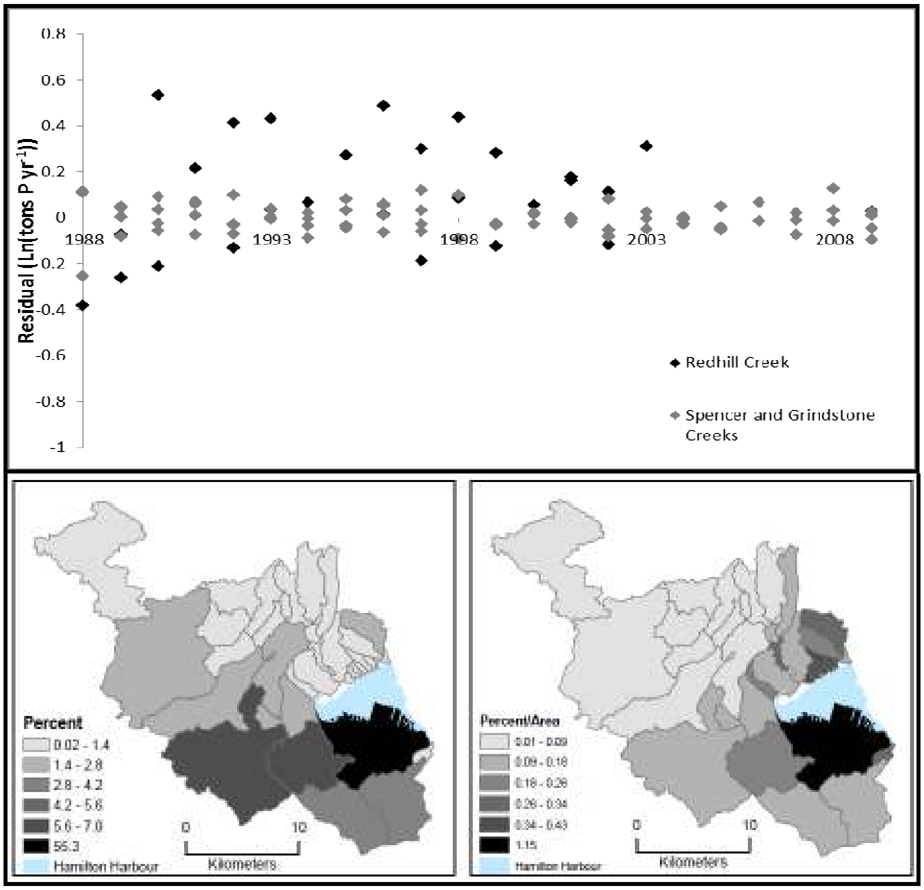

Figure 1: *(Top panel)* Posterior median residuals of the SPARROW model predictions in two major tributaries of the Hamilton Harbour watershed. *(Bottom panel)* Estimated contribution of each subwatershed to the total phosphorus loading in Hamilton Harbour. The map on the left expresses the load of each subwatershed as a percentage of the total phosphorus load, including the combined sewer overflows and taking into account attenuation en route to Hamilton Harbour. The map on the right normalizes the percentage contribution by the corresponding subwatershed areas.

3.2 Eutrophication modelling

A complex eutrophication model was developed that considers the interplay among the following state variables in the epilimnion and hypolimnion of the Hamilton Harbour: nitrate (*NO_3*), ammonium (*NH_4*), phosphate (*PO_4*), generic phytoplankton, cyanobacteria-like phytoplankton, zooplankton, organic nitrogen (*ON*) and organic phosphorus (*OP*). The model was forced with the *SPARROW* outputs. To address the mismatch between the annual predictions of the watershed model and the daily resolution of the model for the receiving waterbody, we developed a Bayesian hierarchical downscaling algorithm. This approach connects the daily precipitation in the watershed with the downstream flows using logistic regression modeling and Bernoulli distribution to reproduce low and high flow regimes. A Bayesian calibration framework was then implemented, founded upon a statistical formulation that

explicitly accommodates measurement error, parameter uncertainty, and model structure imperfection [Ramin *et al.*, 2011].
The model achieved a good representation of several key water quality variables (chlorophyll *a*, total zooplankton biomass, phosphate, and total phosphorus) and sufficiently reproduced the major cause-effect relationships underlying the Harbour dynamics. In particular, the model predicts a weakly positive *Chla-TP* relationship under the present loading conditions, while the corresponding chlorophyll *a* predictive distributions for different *TP* levels consistently exceed the targeted level of 10 μg L^{-1} (Fig. 2a). When the model is forced with the Hamilton Harbour *RAP* nutrient loading propositions, the epilimnetic *TP* concentrations dramatically decrease (< 24 μg L^{-1}), while *TP* levels lower than 20 μg L^{-1} significantly decrease the exceedance frequency of the 10 μg L^{-1} chl *a* goal (Fig. 2b). Further, the relatively discontinuous drop of the chlorophyll *a* predictive distributions around the level of 20 μg *TP* L^{-1} implies a severe accentuation of the phosphorus limitation of the algal growth in the system, given the posterior phytoplankton parameterization obtained. The third panel of the same figure illustrates the predictive distributions of chlorophyll *a* and epilimnetic *TP* concentrations. Generally, the modeling analysis provides evidence that the two criteria are achievable, but the water quality setting process should accommodate the natural variability by allowing for a realistic percentage of violations, e.g., exceedances of less than 10% of the weekly samples during the stratified period should still be considered as system compliance.

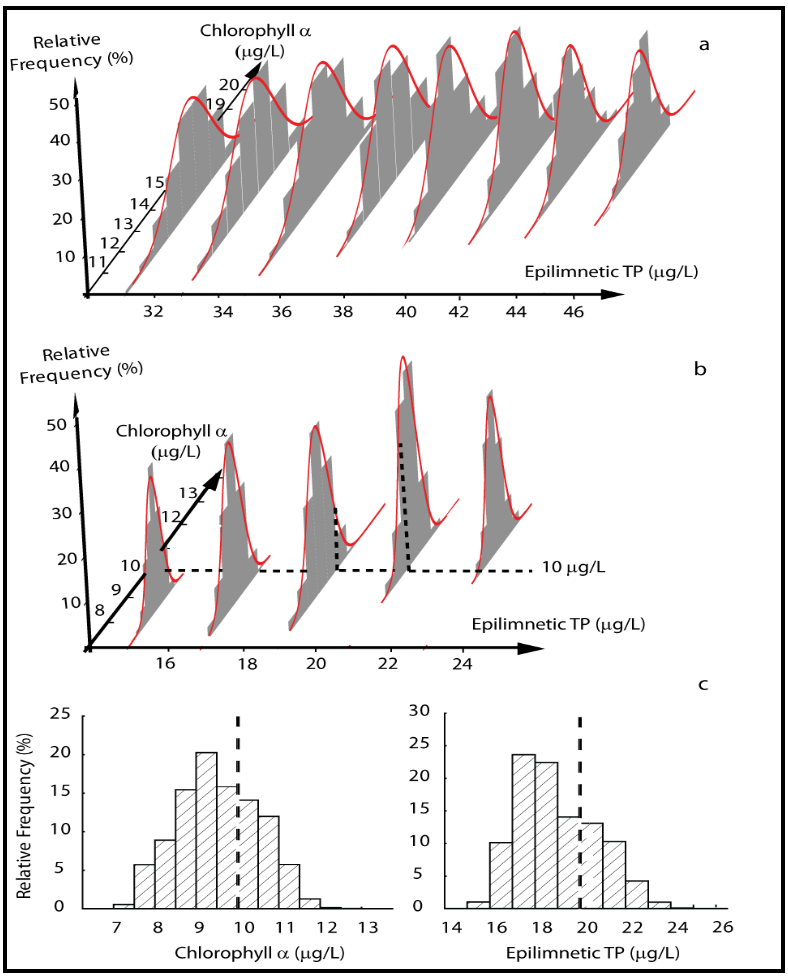

Figure 2: Chlorophyll *a* predictive distributions for different levels of TP concentrations under (a) the present and (b) the Hamilton Harbour RAP loading targets. The third panel (c) illustrates the predictive distributions of chlorophyll *a* and epilimnetic *TP* concentrations derived from the complex eutrophication model.

3.3 Bayesian Model Averaging

Recognizing that there is no true model of an ecological system, but rather several adequate descriptions of different conceptual basis and structure, Bayesian Model Averaging (BMA) is a technique designed to explicitly account for the uncertainty inherent in the model selection process [Raftery *et al.*, 2005]. By averaging over many different competing models, BMA incorporates the uncertainty about the optimal model for any given exercise into the inference drawn about parameters and prediction. Therefore, rather than picking the single “best-fit” model to predict future system responses, we can use Bayesian model averaging to provide a weighted average of the forecasts from different models. In this regard, the projections of the complex eutrophication model were tested against those from a simple model that considers the interplay among the limiting nutrient (phosphate), phytoplankton, zooplankton, and detritus (particulate phosphorus); also known as NPZD model in the literature (Ramin et al., 2012).
The two models represent both ends of the complexity spectrum, characterized by different strengths and weaknesses. One model is a simple mathematical description of the system that accounts for the interplay between the limiting nutrient and aggregated biotic compartments such as “phytoplankton”, and “zooplankton”. This simple approach is more easily subjected to detailed uncertainty analysis and also has the advantage of fewer unconstrained parameters. The second model simulates two elemental cycles, functional phytoplankton groups, and dynamic nutrient release from the sediments. The sophisticated parameterization of the complex model provides confidence for more realistic reproduction of natural system dynamics, but the main criticism for this strategy is the inevitably poor identifiability with respect to the available data as well as the limited flexibility (high computational demands) to thoroughly examine model uncertainty to the input requirements.

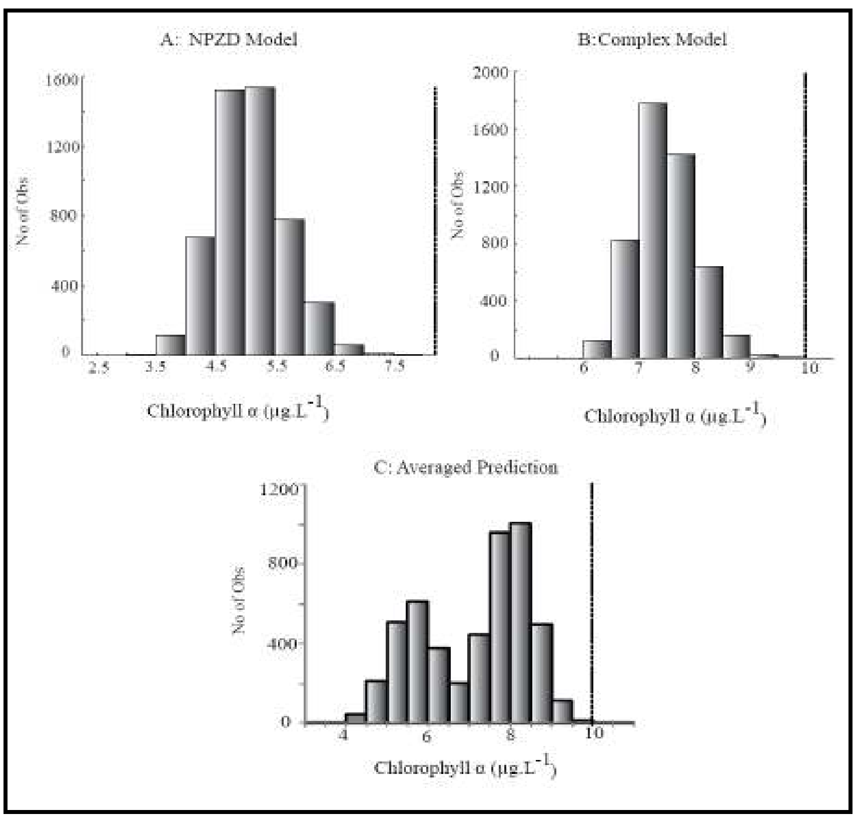

Figure 3: Predictions of the epilimnetic summer chlorophyll *a* concentrations, under the proposed nutrient loading reductions by the Hamilton Harbour RAP, based on the two eutrophication models (*A-B*) and their averaged predictions (*C*).

The predictions from the two models were combined using the respective mean model standard error estimates as weights in a weighted model average:

$$w_{ij} = \frac{\sum_{k=1}^{MC} \frac{\sigma_{ijk}}{\overline{Y}_j}}{MC} \quad \textbf{(2)}$$

$$w_{Mi} = \frac{m}{\sum_{j=1}^{m} w_{ij}} \quad \textbf{(3)}$$

$$\overline{TP} = \sum_{i=1}^{l} w_{Mi} TP_{Mi} \qquad \overline{chla} = \sum_{i=1}^{l} w_{Mi} chla_{Mi} \quad \textbf{(4)}$$

where l represents the number of models considered in this analysis (l = 2); m corresponds to the number of state variables j of the model M_i for which data are available (m = 6 or 11); MC is the total number of MCMC runs sampled to form the model posteriors; σ_{ijk} denotes the model structural error for the state variable j of the model M_i as sampled from the MCMC run k; $\overline{Y}_j$ represents the annual observed average for the variable j, TP_{Mi} and $chla_{Mi}$ are the total phosphorus and chlorophyll a predictions from the individual models weighted by the corresponding weights w_{Mi} to obtain the averaged predictions $\overline{TP}$ and $\overline{chla}$.

In particular, both models also predict that the epilimnetic chlorophyll *a* concentrations will fall below the threshold level of 10 μg *chla* L^{-1} (Fig. 3). Yet, the simple model appears to support more optimistic predictions with respect to phytoplankton response to the reduced ambient *TP* concentrations relative to the complex one. Consequently, the averaged predictive distribution for chlorophyll *a* demonstrates a distinct bimodal pattern with a primary mode at 7.5 μg *chla* L^{-1}, reflecting the greater weight (higher performance) of the complex model, and a secondary peak at 5.1 μg *chla* L^{-1}, associated with the simple one (Fig. 3). One of the major structural differences of the two models lies in the way they handle the nutrient fluxes from the sediments, i.e., a static phosphorus flux vis-à-vis a mechanistic characterization that relates phosphorus release to particulate sedimentation and burial rates [Ramin *et al.*, 2011]. Being part of the model updating process, the simple model predicts that the sediments contribute approximately 1.1 mg *P* m^2 day^{-1} into the overlying water column, whereas the same fluxes are raised up to 2.0 mg *P* m^2 day^{-1} with the complex model. Under the reduced nutrient loading scenario, the dynamic nature of the sediment response with the complex model decreases the release of phosphorus at the level of 1.5 mg m^2 day^{-1}, which however remains well above the flux used to force the simple model. This discrepancy most likely reflects one of its structural weaknesses and also highlights the importance of embracing more sophisticated approaches to sediment diagenesis in the Harbour. Despite all the arguments historically used to downplay the relative contribution of the sediment fluxes in the system, recent evidence suggests that the hypolimnetic phosphate can easily exceed the level of 30 μg PO_4 L^{-1} for extended period (3-4 weeks) during the late summer/early fall (T. Labencki, unpublished data). This pattern likely suggests that the summer epilimnetic environment may also be subject to intermittent nutrient pulses from the hypolimnion, which in turn can have profound ramifications on the dynamics of the phytoplankton community.

4. Discussion-Future Perspectives

Modellers must acknowledge the uncertainty pertaining to the selection of the optimal model structure for a specific environmental management problem, and Bayesian averaging of two or more models is a promising means for improving the contemporary modelling practice. In the context of ecological process-based modelling though, this approach should not be viewed solely as a framework to improve our predictive devices, but rather as an opportunity to compare alternative ecological structures, to challenge existing ecosystem conceptualizations, and to integrate across different (and often conflicting) paradigms. Future research should also focus on the refinement of the weighting schemes and other performance standards to impartially synthesize the predictions of different models. Several interesting statistical post-processing methods presented in the field of ensemble weather forecasting will greatly benefit our attempts to develop weighting schemes suitable for the synthesis of multiple ecosystem models. Some of the outstanding challenges involve the development of ground rules for the features of the calibration and validation domain [Anderson, 2005], the inclusion of penalties for model complexity that will allow building forecasts upon parsimonious models, and performance assessment that does not exclusively consider model endpoints but also examines the plausibility of the underlying ecosystem structures, i.e., biological rates, ecological processes or derived quantities [Arhonditsis and Brett, 2004].

In conclusion, Bayesian inference techniques are uniquely suitable for integrating various types of models (complex dynamic models, empirical equations, expert judgments) into one coherent framework, while rigorously assessing the uncertainty associated with model structures, parameters and other inputs. In particular, my recent research has shown that the Bayesian paradigm can effectively alleviate problems of spatiotemporal resolution mismatch among different submodels of integrated environmental modelling systems, overcome the conceptual or scale misalignment between processes of interest and supporting information, exploit disparate sources of information that differ with regards to the measurement error and resolution, and accommodate tightly intertwined environmental processes operating at different spatiotemporal scales

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.

REFERENCES

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

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

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

Borsuk, M.E., Stow, C.A., and Reckhow, K.H.. Predicting the frequency of water quality standard violations: a probabilistic approach for TMDL development. *Environ. Sci. Technol*. 36(10):2109–2115, 2002.

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

Raftery, A.E., Balabdaoui, F., Gneiting, T., and Polakowski, M., Using Bayesian model averaging to calibrate forecast ensembles, Technical Report 440, Department of Statistics, University of Washington, Seattle, 2003.

Ramin, M., S. Stremilov, T. Labencki, A. Gudimov, D. Boyd, and G.B. Arhonditsis, Setting water quality criteria in Hamilton Harbour, Ontario, Canada: A Bayesian approach. *Environmental Modelling & Software* 26, 337-353, 2011.

Ramin M., T. Labencki, D. Boyd, D. Trolle, and G.B. Arhonditsis. A Bayesian synthesis of predictions from different models for setting water quality criteria. *Environmental Modelling & Software, In Press.*

Wellen, C., G.B. Arhonditsis, T. Labencki, and D. Boyd, A Bayesian methodological framework to accommodate inter-annual nutrient loading variability with the SPARROW model. *Water Resources Research In Press.*