Environmental Science & Technology

Article

pubs.acs.org/est

# Continuous Bayesian Network for Studying the Causal Links between Phosphorus Loading and Plankton Patterns in Lake Simcoe, Ontario, Canada

Alexey Gudimov,[†] Eavan O'Connor,[‡] Maria Dittrich,[†] Hamdi Jarjanazi,[§] Michelle E. Palmer,[§] Eleanor Stainsby,[§] Jennifer G. Winter,[§] Joelle D. Young,[§] and George B. Arhonditsis*,[†]

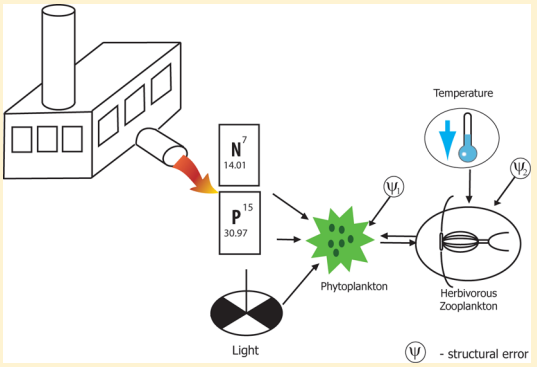
[†]Ecological Modeling Laboratory, Department of Physical and Environmental Sciences, University of Toronto, Toronto, Ontario, Canada, M1C 1A4

[‡]Lake Simcoe Region Conservation Authority, Newmarket, Ontario, Canada, L3Y 4X1

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

Supporting Information

**ABSTRACT:** An ecosystem perspective to restoring beneficial uses in Areas of Concern can be interpreted as a shift from the traditional elucidation of simple cause–effect relationships to a multicausal way of thinking that more effectively accommodates ecosystem complexity. This holistic management paradigm has also pervaded the contemporary ecological modeling practice, making compelling the adoption of more sophisticated ecosystem modeling tools. In this study, our primary objective is to develop a Bayesian hierarchical network of simple ecological models for Lake Simcoe, Ontario, Canada, aiming to establish a realistic representation of the causal connections among exogenous nutrient loading, ambient nutrient conditions, and epilimnetic plankton dynamics. In particular, we used a spatially explicit simple mass-balance model forced with idealized sinusoidal loading to predict total phosphorus concentrations. A structural equation model was then used to delineate the interplay among nutrients, ambient light conditions, phytoplankton, and herbivorous biomass. Our analysis highlights the strength of the causal linkages between total phosphorus and water clarity with phytoplankton as well as the capacity of zooplankton grazing to modulate the algal standing crop. Our Bayesian network is also used to examine the exceedance frequency of threshold values for total phosphorus (15 $\mu$g/L) and chlorophyll *a* (4 $\mu$g/L) concentrations under scenarios of phosphorus loading reduction. Our study suggests that a 15% phosphorus loading decrease will still result in >25% violations of the 4 $\mu$g chl*a*/L value in the two embayments of Lake Simcoe (Cook's Bay and Kempenfelt Bay). The TP levels will decrease in response to the exogenous loading reductions and this improvement will be primarily manifested in the northcentral segments of the system.




## 1. INTRODUCTION

Addressing environmental management problems often involves complex policy decisions that aim to maintain ecosystem functional integrity while accommodating social values and economic concerns. The growing appreciation of the challenges of aquatic ecosystem restoration and the need to address a wide array of tightly intertwined stressors have triggered a shift from the historical water quality/fisheries exploitation paradigms to the ecosystem management paradigm.[1] Rather than narrowly focusing on water quality problems, the ecosystem approach simultaneously addresses problems related to fisheries management, exotic species invasions, biodiversity, habitat conservation and restoration, sustainable economic development, or even human behavior and education.[2] Although the concept of holistic ecosystem management makes sense as a pragmatic means to address multifaceted environmental problems, there is concern that this approach has been accompanied by a shift from the traditional elucidation of simple cause–effect relationships to a multicausal way of thinking to accommodate ecosystem complexity.[1] This shift can be a major impediment to deriving the straightforward scientific answers required by the regulatory agencies tasked with implementing environmental protection policies.[3]

The ecosystem approach has not only pervaded environmental thinking but also contemporary modeling practices. Complex ecosystem models have been developed for a number of purposes, including the illumination of causal mechanisms, complex interrelationships, and direct and indirect ecological paths; examination of the interplay among a suite of external

Received: March 14, 2012
Revised: May 27, 2012
Accepted: June 5, 2012
Published: June 5, 2012

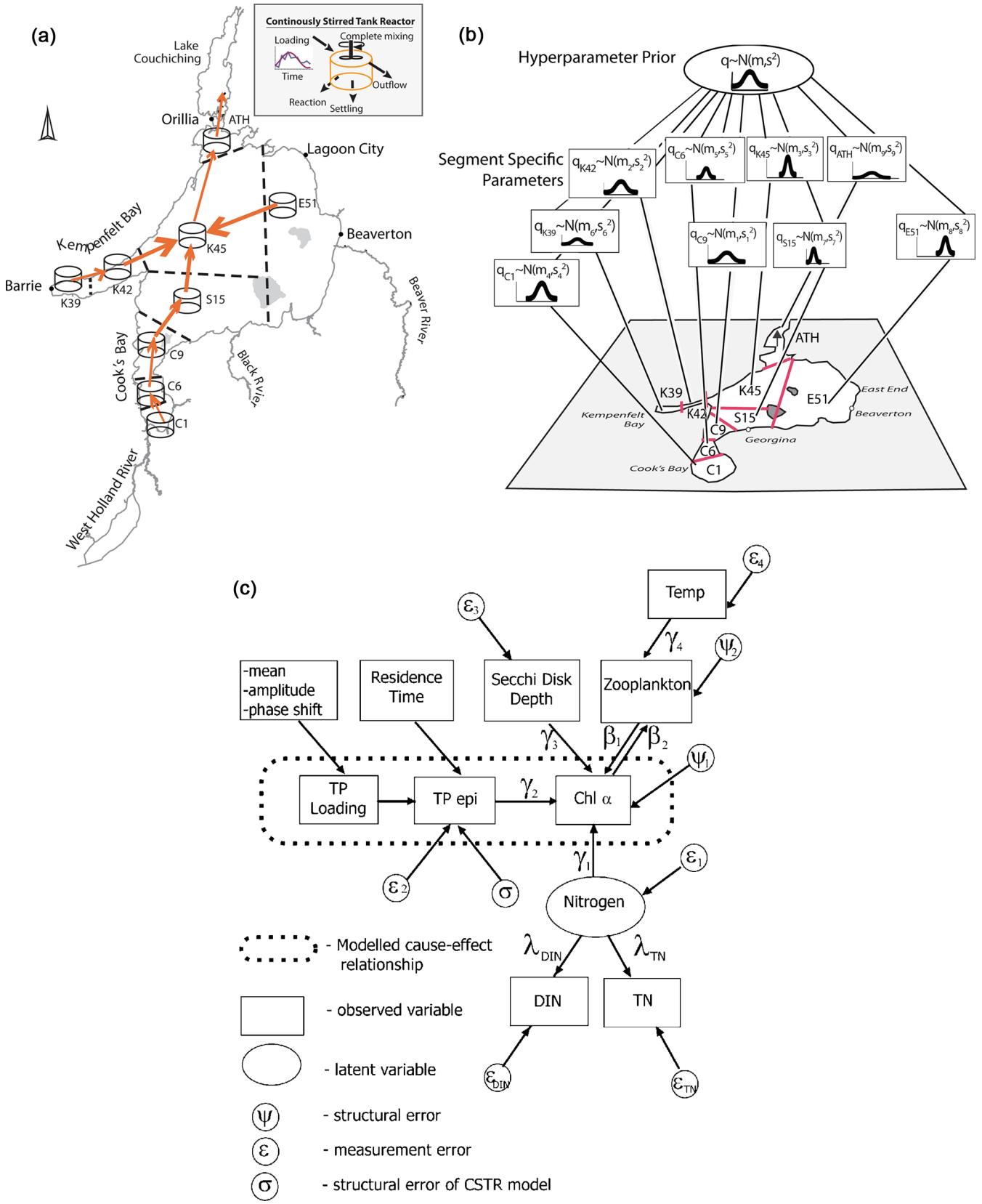



**Figure 1.** (A) Basic principle and spatial segmentation of the Continuously-Stirred-Tank Reactor (CSTR) model, forced with idealized sinusoidal loading, to predict epilimnetic total phosphorus concentrations in Lake Simcoe, Ontario, Canada. (B) Hierarchical configuration of the Lake Simcoe Bayesian network. The spatial heterogeneity is accommodated by viewing the problem of parameter estimation as a hierarchy. At the bottom of the hierarchy are the parameters for individual segments $j$; $q_j \sim N(m_j, s_j)$. In the upper level, the segment-specific parameters are specified probabilistically in terms of lakewide parameters or hyper-parameters; $q \sim N(m, s)$. (C) Structural equation model used to elucidate the interplay among nutrients, ambient light conditions, phytoplankton, and herbivore biomass.


dx.doi.org/10.1021/es300983r | Environ. Sci. Technol. 2012, 46, 7283−7292

stressors (e.g., climate change, urbanization/land-use changes, invasion of exotic organisms); and assessment of the potential ramifications of various stressors on ecosystem functioning (e.g., food web dynamics, benthic–pelagic coupling, fish communities).[4,5] While the development of more holistic modeling constructs is certainly the way forward, skeptics question the ability to mathematically depict many biotic relationships and their interactions with the abiotic environment.[6,7] The current generation of mathematical models cannot offer robust predictions in a wide range of spatiotemporal domains, and our experience is that model performance declines as we move from physical–chemical to biological components of aquatic ecosystems.[6] The tendency to invoke complexity as a means for improving the learning capacity of our models also entails an increase in the disparity between what ideally we want to learn (internal description of the system and model end points) and what can realistically be observed. Thus, our ability to properly constrain model parameters from observations is reduced, and the resulting poor identifiability undermines model credibility when used to support environmental management decisions.[8]

In this study, our thesis is that while the emergence of the holistic environmental management paradigm reinforces the need for more sophisticated modeling tools, the critical evaluation of the inference drawn and the impartial differentiation between real knowledge gained and existing knowledge gaps can be the thrusts for coping with the uncertainty of any modeling exercise. We also advocate the use of simple models as adequate first-order approximations until simplicity can be gradually traded for increased explanatory power. In this regard, our primary objective is to develop a continuous Bayesian hierarchical network of simple ecological models for Lake Simcoe, Ontario, Canada, that establish a realistic representation of the causal connections among exogenous nutrient loading, ambient nutrient conditions, and epilimnetic plankton dynamics. We use a spatially explicit simple mass-balance model forced with idealized sinusoidal loading to predict total phosphorus concentrations. A structural equation model is then used to delineate the interplay among nutrients, ambient light conditions, phytoplankton, and herbivorous zooplankton biomass. Our study also pinpoints directions of future model augmentation, where extra ecological complexity and finer spatial segmentation will improve our understanding of the system. Finally, we illustrate how our Bayesian network of models can be used to examine the exceedance frequency of threshold values for total phosphorus (15 $\mu$g/L) and chlorophyll *a* (4 $\mu$g/L) concentrations under scenarios of phosphorus loading reduction.

## 2. METHODS

**2.1. Study Site.** Lake Simcoe is the sixth largest inland lake in the province of Ontario, Canada, with a surface area of 722 km$^2$, a mean depth of 14 m, and a maximum depth of 42 m (Figure 1a). It is a dimictic system that completely freezes over in most winters. Lake Simcoe consists of a large main basin (mean depth 14 m, maximum depth 33 m) and two large bays: the narrow and deep Kempenfelt Bay on the west side of the lake (area 34 km$^2$, mean depth 20 m) and the shallow Cook's Bay at the south end of the lake (area 44 km$^2$, mean depth 13 m).[9] The lake drains through a single outflow at Atherley Narrows and has a flushing time of approximately 11 years.[10] Due to the limestone bedrock underlying its catchment, Lake Simcoe is a hard-water lake with mean calcium concentration of 41 mg/L, mean alkalinity of 116 mg/L, and mean sulfate concentration of 20 mg/L.[10] The lake supports a year-round sport fish industry (>1 million angler hours per year) as well as recreational activities that generate over $200 million per year. Lake Simcoe is also a drinking water source for several communities within its 2899 km$^2$ watershed.[9]

Agriculture and increasing urbanization activities have impacted the ecological health of the system. In particular, Lake Simcoe currently receives wastewater from 14 municipal wastewater treatment plants, which constitute sources of phosphorus loading (6 $\pm$ 1 tonnes/yr between 2004 and 2007).[10] Substantial phosphorus loads are also deposited from the atmosphere (18 $\pm$ 4 tonnes/yr) or stem from other nonpoint sources, including runoff from agricultural, urban, and natural areas (43 $\pm$ 5 tonnes/yr) and rural septic systems (4.4 $\pm$ 0.1 tonnes/yr).[10] The exogenous phosphorus inputs modulate the ambient total phosphorus (TP) levels and subsequently trigger phytoplankton production,[11] while the decomposition of the excessive organic material in the sediments likely contributes to hypolimnetic dissolved oxygen (DO) depletion. Prior to the mid-1990s, end-of-summer hypolimnetic DO values reached nearly lethal levels for many coldwater fish species (<3 mg/L).[12] As a result, fish biomass declined for several commercially or recreationally important fish species, such as lake trout (*Salvelinus namaycush*), lake whitefish (*Coregonus clupeaformis*), and lake herring (*Coregonus artedi*).[12] To alleviate the problem of hypoxia and thus allow for the restoration of a self-sustaining coldwater fishery, the target for the end-of-summer minimum volume-weighted hypolimnetic dissolved oxygen (MVWHDO) was originally established at 5 mg/L and recently revised to 7 mg/L.[9,12] A combination of empirical knowledge and modeling indicates a phosphorus loading rate of 44 tonnes/yr is needed to meet the 7 mg/L MVWHDO target.[9] Between 2004 and 2007, total phosphorus loading into the lake was 74 $\pm$ 3 tonnes/yr, a significant reduction from over 100 tonnes/yr during the 1980s and early 1990s.[12]

**2.2. Model Description.** *2.2.1. Spatial Segmentation.* Our continuous Bayesian network consists of two models: (i) a spatially explicit simple mass-balance model forced with idealized sinusoidal loading to predict epilimnetic total phosphorus; and (ii) a structural equation model to depict the causal links among nutrients, light availability, temperature variability, phytoplankton, and herbivorous zooplankton biomass in the Lake Simcoe epilimnion. Data for the model parameterization were obtained from the monitoring program conducted by the Ontario Ministry of the Environment in partnership with the Lake Simcoe Region Conservation Authority. Nine stations (C1, C6, C9, K39, K42, K45, S15, E51, and ATH) have been historically used to monitor the water quality in Lake Simcoe and these stations were also used to delineate the spatial segmentation of the model (Figure 1a). Loading estimates from all the major point and nonpoint sources during the study period (1999–2007) were provided by the Lake Simcoe Region Conservation Authority.[13] (More information about the data used in the present study is provided in the Supporting Information). Our modeling analysis is based on a hierarchical structure,[14] which allowed us to obtain segment-specific estimates of the causal relationships considered to accommodate the spatial variability in the system (Figure 1b).

*2.2.2. Total Phosphorus Model.* Our TP model is founded upon a representation of the Lake Simcoe epilimnion as a feedforward system of completely mixed reactors (Figure 1a).[15] A central feature of the feedforward configuration is the postulation of a net unidirectional flow within the framework of serial reactors considered. This approach seems conceptually more suitable to model horizontal mass exchanges in a chain of

lakes or a stream,[15] and thus its validity and limitations to accommodate the spatial heterogeneity of a single lake is critically examined in the Supporting Information (Section D). The TP balance in each segment is determined by the exogenous loading sources and three sinks (outflow, reaction, and settling) that deplete the ambient phosphorus levels in the system:

$$\frac{\mathrm{dTP}}{\mathrm{d}t} = \frac{W_{(t)}}{V} - \frac{Q}{V}\cdot\mathrm{TP} - k\cdot\mathrm{TP} - \frac{\nu}{H}\cdot\mathrm{TP} \quad (1)$$

where $W_{(t)}$ denotes the loading entering the segment (tonnes/day), $Q$ refers to the volumetric outflow rate for the segment ($m^3$/day), $V$ is the volume of the segment ($m^3$), $\nu$ represents the settling velocity (m/day), $H$ is the mean depth of the segment (m), and $k$ denotes the first-order reaction coefficient ($day^{-1}$). Under the assumption that the seasonal exogenous TP loading follows a sinusoidal pattern, the solution of the differential equation that describes the temporal variability of the epilimnetic TP is

$$\mathrm{TP} = \frac{W_{\mathrm{avg}}}{\lambda V} + \frac{W_{\mathrm{amp}}}{V\sqrt{\lambda^2+\omega^2}}\sin[\omega t - \theta - \phi(\omega)]$$

$$\omega = \frac{2\pi}{T},\ \phi(\omega) = \arctan\left(\frac{\omega}{\lambda}\right)$$

$$\lambda = \kappa_1 + \kappa_2,\ \kappa_1 = \frac{Q}{V},\ \kappa_2 = k + \frac{\nu}{H} \quad (2)$$

in which $W_{\mathrm{avg}}$ is the mean loading entering the system (tonnes $day^{-1}$), $W_{\mathrm{amp}}$ is the amplitude around the mean loading (tonnes $day^{-1}$), $\theta$ = phase shift of the loading from the standard wave (radians), $\varphi(\omega)$ is an additional phase shift related to the segment-specific response, and $\omega$ (radians $day^{-1}$) and $T$ (day) are the angular frequency and period of the loading oscillation. The hierarchical formulation used to accommodate the spatiotemporal variability of the TP concentrations was specified as follows:

$$\log(\mathrm{TP}_{itj}) \sim N(f(\kappa_{1tj}, \kappa_{2tj}, W_{\mathrm{avg}tj}, W_{\mathrm{amp}tj}, \theta_{tj}), \tau^2)$$

$$1/\tau^2 \sim G(0.001, 0.001)$$

$$\kappa_{1tj} \sim N(\kappa_{1j}, {\sigma_{\kappa 1j}}^2)$$

$$\kappa_{2tj} \sim N(\kappa_2, {\sigma_{\kappa 2j}}^2),\ \kappa_2 \sim N(\kappa_{2\mu}, {\sigma_{\kappa 2}}^2)$$

$$\kappa_{2\mu} \sim N(\kappa_{2\mu\mathrm{lit}}, {\sigma_{\kappa 2\mathrm{lit}}}^2),\ 1/{\sigma_{\kappa 2}}^2 \sim G(a_1, a_2),\ 1/{\sigma_{\kappa 2j}}^2 \sim G(a_{1j}, a_{2j})$$

$$W_{\mathrm{avg}tj} \sim N(W_{\mathrm{avgML}tj}, {\textstyle\sum_{\mathrm{Tot}tj}}^2)$$

$${\textstyle\sum_{\mathrm{Tot}tj}}^2 = {\sigma_{Wtj}}^2 + {\sigma_{W\mathrm{avg}tj}}^2$$

$$W_{\mathrm{amp}tj} \sim N(W_{\mathrm{amp}MLtj}, {\sigma_{\mathrm{amp}tj}}^2)$$

$$\theta_{tj} \sim N(\theta_{\mathrm{ML}tj}, {\sigma_{\theta tj}}^2)$$

$$i = 1, \ldots, 12,\ j = 1, \ldots, 9,\ t = 1, \ldots, 9 \quad (3)$$

where $\mathrm{TP}_{itj}$ corresponds to the average TP concentration at segment $j$, year $t$, and month $i$; $\tau^2$ denotes the structural error variance of our TP model; $\kappa_{1tj}$ is the net outflow rate from segment $j$ at year $t$; the $\kappa_{1j}$ and $\sigma_{\kappa 1j}$ correspondingly denote the segment-specific average flushing rates and the associated interannual variability, as calculated from the respective water balance budgets; $\kappa_{2tj}$ is the net TP loss rate in segment $j$ at year $t$; $\kappa_2$ represents the hyperparameter or the lakewide TP loss rate; ${\sigma_{\kappa 2j}}^2$ is the segment-specific variance; and $\kappa_{2\,\mu}$ and ${\sigma_{\kappa 2}}^2$ are the mean and variance of the global parameter distribution, respectively. The normal distribution assigned to the parameter $\kappa_{2\,\mu}$ was based on a literature review ($\kappa_{2\,\mu\mathrm{lit}}$, ${\sigma_{\kappa 2\mathrm{lit}}}^2$),[14] while the inverse gamma distributions assigned to the parameters ${\sigma_{\kappa 2}}^2$ and ${\sigma_{\kappa 2j}}^2$ were constructed such that their mean was equal to the variance ${\sigma_{\kappa 2\mathrm{lit}}}^2$. The uncertainty of the same distributions reflected our high level of confidence to that mean variance estimate (i.e., coefficient of variation <10%). The mean ($W_{\mathrm{avg}tj}$), amplitude ($W_{\mathrm{amp}tj}$), and phase shift ($\theta_{tj}$) values of the phosphorus loading at segment $j$, year $t$, and month $i$ were drawn from normal distributions in which the mean values ($W_{\mathrm{avgML}tj}$, $W_{\mathrm{ampML}tj}$, $\theta_{\mathrm{ML}tj}$) and error variances (${\sigma_{W\mathrm{avg}tj}}^2$, ${\sigma_{\mathrm{amp}tj}}^2$, ${\sigma_{\theta tj}}^2$) were the maximum likelihood estimators obtained from fitting sinusoidal functions to the corresponding monthly loading data. The normal distributions of the mean annual loading ($W_{\mathrm{avg}tj}$) also considered the estimates of model error variance (${\sigma_{Wtj}}^2$) obtained from the maximum likelihood fitting exercise.

*2.2.3. Structural Equation Model.* We used structural equation modeling (SEM) to elucidate the key causal relationships underlying the interplay among the physical environment, nutrients, and plankton dynamics in Lake Simcoe. SEM is a multivariate statistical method that encompasses both factor and path analysis, which allows decomposing multiple causal pathways and quantifying direct and indirect relationships among variables.[16] Another advantage of SEM is that it can explicitly incorporate uncertainty due to measurement error and/or accommodate the discrepancy between conceptual ecosystem properties and observed variables that can be directly measured. SEM is also an a priori statistical method whereby a hypothetical structure of the system studied, reflecting the best knowledge available, is tested against the observed covariance structure.[16] A Bayesian approach to SEM was also adopted because it offers several advantages over the classical methods (e.g., maximum likelihood, generalized and weighted least-squares). A Bayesian SEM can incorporate prior knowledge about the parameters and more effectively treat unidentified models. Moreover, the modeling process does not rely on asymptotic theory, a feature that is particularly important when the sample size is small and the classical estimation methods are not robust. Markov Chain Monte Carlo (MCMC) samples are taken from the posterior distribution, and consequently the procedure works for all sample sizes and various sources of non-normality.[17,18]

In this study, our starting point is a "conceptual/mental" model that considers the effects of four latent variables (i.e., phosphorus, nitrogen, light, and herbivorous zooplankton grazing) on phytoplankton dynamics (Supporting Information Figure SI-1). Each of these conceptual factors (or latent variables) can be linked with observed data (rectangular boxes in Figure 1c), while explicitly acknowledging that none of the selected surrogate variables can perfectly represent the underlying property (measurement errors $\varepsilon$ in Figure 1c). Specifically, we established a connection between the TP mass-balance and structural equation models by considering the causal association between total phosphorus and phytoplankton biomass. We also used the latent variable nitrogen and two indicator variables: dissolved inorganic nitrogen (DIN) and total nitrogen concentrations. The role of light availability (light limitations on algal growth) was described solely by the existing Secchi disk

depth values, while chlorophyll *a* concentrations were used to represent the latent variable "phytoplankton". The trophic interactions between phytoplankton community and herbivorous zooplankton were assumed to have a nonrecursive nature, i.e., bottom-up and top-down forcing. The role of temperature as a regulatory factor of zooplankton biomass was also considered. The Bayesian hierarchical configuration of the Lake Simcoe SEM is provided in Supporting Information, Section D.

## 3. RESULTS

The application of the TP model provided satisfactory fit to the measured TP concentrations in seven out of the nine segments considered, resulting in root-mean-square-error (RMSE) values lower than 5 $\mu$g TP/L (Figure SI-2). The main exceptions were the two inner segments of Cook's Bay, C1 (RMSE = 26.47 $\mu$g TP/L) and C6 (RMSE = 10.77 $\mu$g TP/L), which are also consistently characterized by the highest TP levels in the system. In particular, we note that the fairly high error in segment C1 predominantly stems from a few extreme TP values that occurred distinctly out of the phase shifts, $\varphi(\omega_{jt})$, postulated by our model to account for the system response. Further, one of the fundamental model assumptions is that the segment-specific TP seasonal patterns are mainly driven by the intra-annual variability of the exogenous loading. In this regard, our model predicts that the inflowing TP loads alone can induce oscillatory behavior in Cook's Bay, although the agreement between actual and predicted amplitudes cannot be rigorously verified as the typical sampling intensity (biweekly) and timing (mid-May) may not always coincide with the observed TP peaks in Lake Simcoe (Figure SI-2). On the other hand, the predicted TP dynamics in larger and/or offshore segments were significantly muted, indicative of a more complex interplay among exogenous loading, hydrodynamics, and biological productivity that most likely modulates in-lake TP variability.

In the majority of the modeled sites, the net TP loss rates were well-identified and independent from the corresponding posterior estimates of the segment-specific flushing (or net exchange) rates (Figure SI-3). They demonstrated low year-to-year variability within each segment, spanning a range between 0.8 and 1.0 year$^{-1}$ or 0.002−0.003 day$^{-1}$ (Figure SI-4 and Table SI-1). Notable exceptions were the segments C6 and K39, characterized by 2- to 10-fold increase in their posterior $\kappa_2$ estimates with considerable interannual variability. Given that both sites are located in the vicinity of major tributary outlets, there are three plausible explanations for this substantial discrepancy: (i) the excessive macrophyte growth in the favorable environment of Cook's Bay (i.e., shallower depths, elevated nutrient levels, increased water clarity, and fine-grained sediments) that act as a net sink of the ambient TP;[19] (ii) the presence of dreissenids that have the capacity to filter suspended particles from the water column and thus modulate the TP concentrations;[20] and/or (iii) the fact that both segments constitute transitional zones to the deeper parts of the lake. It is conceivable that a substantial portion of the inflowing TP loads may be directly allotted to the hypolimnion, representing a permanent sink for the epilimnetic phosphorus budget (see also following discussion). Generally, the posterior estimates of the sedimentation and outflow rates suggest that a significant portion of the annual TP inputs from the Holland River in the southernmost segment (C1) are flushed into the middle area (C6) of Cook's Bay, i.e., 1140 out of 1201 mg/m$^2$/yr (Figure 2 and Figure SI-5). As previously mentioned, a substantial net amount of phosphorus is lost (976 mg/m$^2$/yr) in the sediments

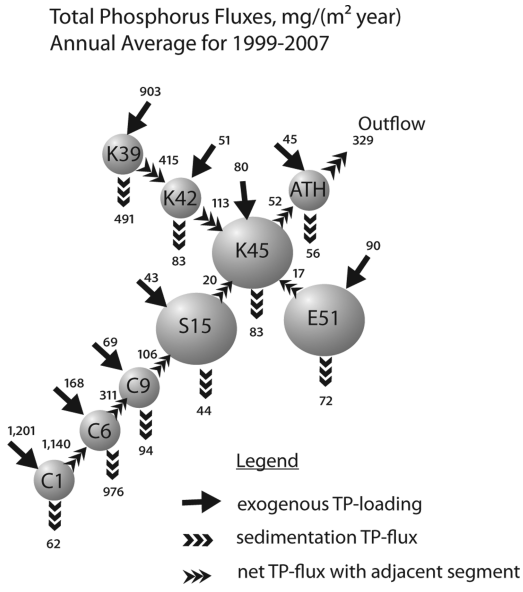



**Figure 2.** Predicted annual phosphorus mass balance at the different segments of Lake Simcoe during the study period (1999−2007).

of this zone or in the hypolimnion of the next segment (C9), while a lower fraction (311 mg/m$^2$/yr) is transported horizontally to the epilimnion of the outer Bay and subsequently to the Main Basin (106 mg/m$^2$/yr). In a similar manner, more than half of the TP loads (or 491 mg/m$^2$/yr out of 903 mg/m$^2$/yr) in the innermost segment of Kempenfelt Bay are being subjected to sedimentation and the remaining amount of phosphorus (415 mg/m$^2$/yr) is being transferred to the outer segment (K42), ultimately reaching the central area of the lake (113 mg/m$^2$/yr). Notably, the net areal sedimentation rates in the mouths of the two embayments were comparable with those predicted in the Main Basin, ranging from 45 to 90 mg/m$^2$/yr. The annual TP outflows from Atherley Narrows were estimated at an average level of 329 mg/m$^2$/yr (or 9.6 tonnes/yr), and the TP retention fraction varied from 85% to 93%; both values are on par with existing estimates independently calculated from TP mass balance budgets for Lake Simcoe.[10,20]

The comparison between measured and predicted chlorophyll *a* concentrations is presented in Figure SI-6, in which it can be seen that our SEM approach sufficiently describes the observed chlorophyll *a* patterns as nearly all the data were included within the 95% credible intervals. The RMSE values were generally lower than 2 $\mu$g chl*a*/L, except in the southernmost segment in Cook's Bay (C1 $\approx$ 4.8 $\mu$g chl*a*/L). In a similar manner, the observed variability of the herbivorous zooplankton biomass is adequately depicted by our model (RMSE < 60 $\mu$g C/L). The largest discrepancy was found in the outer part of Cook's Bay (C6, C9) and the southern Main Basin (S15), where the error was greater than 70 $\mu$g C/L and mainly reflected the influence of several outlying observations. Our SEM analysis primarily highlights the importance of a positive causal link between epilimnetic TP and phytoplankton biomass (chlorophyll *a*); especially, in Kempenfelt Bay and Eastern Basin (Figure 3 and Table SI-2). Interestingly, our model predicts a significant path from nitrogen to phytoplankton in several segments of the lake (C6, C9, K42, K45), and the negative nature of this relationship stems from the inclusion of dissolved phase inorganic nitrogen in the latent variable "nitrogen".[17,18] However, we do not believe the statistically significant signature implies nitrogen limitation in

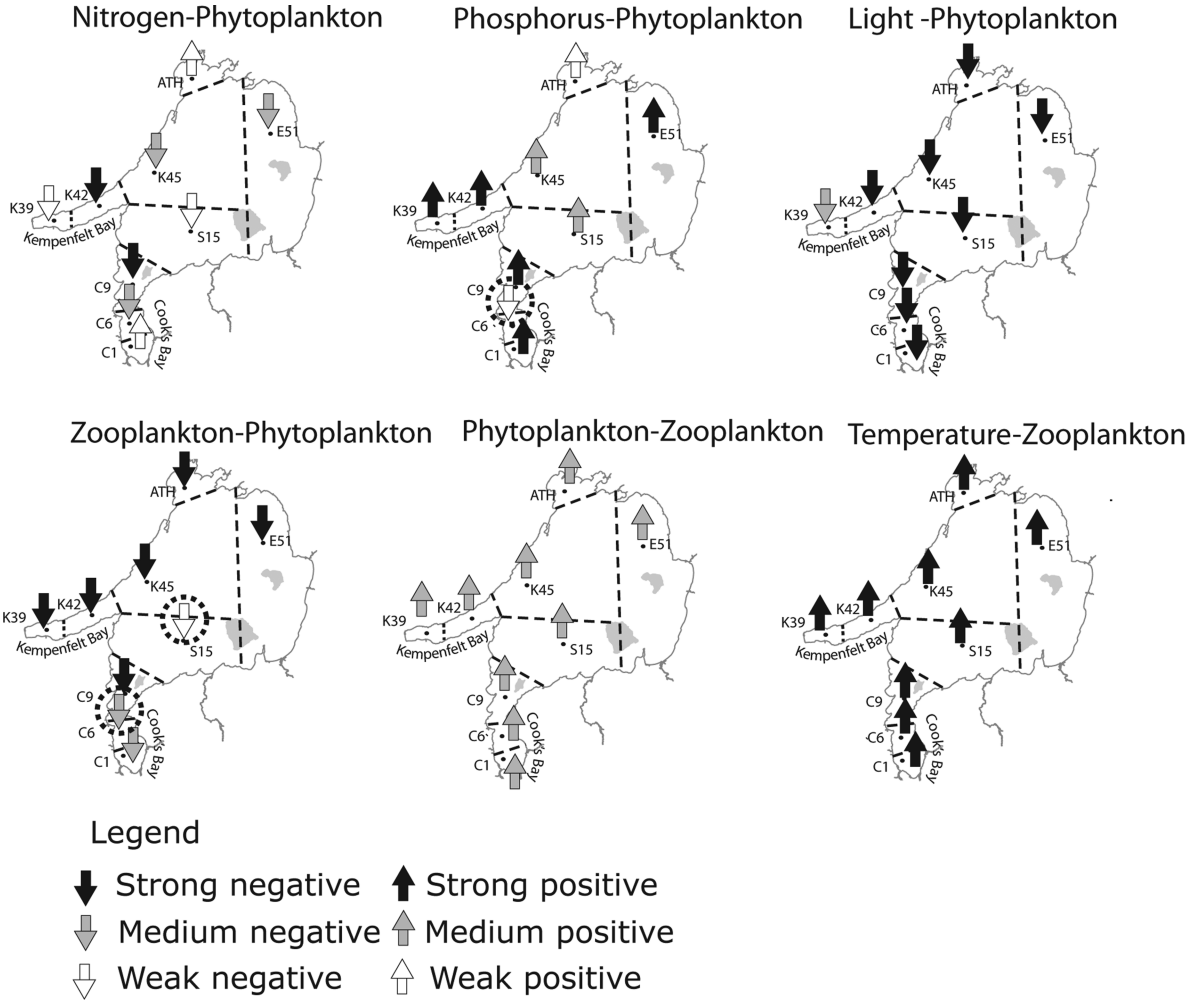



**Figure 3.** Structural paths underlying plankton patterns in the nine segments of Lake Simcoe. The direction and color of the arrows represent the magnitude and sign of the posterior medians of the standardized path coefficients. The standardized coefficients correspond to the shift in standard deviation units of the dependent variable that is induced by shifts of one standard deviation units in the explanatory variables, and thus provide a means to assess the relative importance of the various model paths.

the system. Ambient dissolved phase nutrient concentrations partly reflect the residual nutrients from phytoplankton activity, but there are several other potentially influential factors that are not accounted for in our approach (e.g., intracellular storage). A strongly negative relationship exists between Secchi disk transparency and phytoplankton biomass throughout the lake. The capacity of zooplankton grazing to modulate the algal standing crop (zooplankton→phytoplankton) is consistently manifested in the lake, although this trophic linkage was somewhat weaker and poorly identified (or highly uncertain) in the segments associated with the highest predictive error for zooplankton biomass (C6, S15). The causal path from phytoplankton to zooplankton (bottom-up forcing) was consistently weak. Water temperature variability appears to be a significant driver of zooplankton abundance in the system (Figure 3).

The MCMC posterior samples were used to update the two models and subsequently integrate them into one coherent framework for examining the exceedance frequency of different water quality standards. Namely, the updated mass-balance model was used to derive the predictive TP distributions, stemming from both input uncertainty and model error, which were then propagated through the SEM. To predict the effect of a substantial reduction in phosphorus inputs, the corresponding marginal TP loading distributions were reduced by 15% and 30%. All other functions and marginal nodes in the models were left unchanged, and new distributions were computed for the ecological variables of interest. For illustration purposes, we selected two threshold levels for TP (15 $\mu$g/L) and chlorophyll *a* (4 $\mu$g/L) concentration, representing the midpoint and the lower bound of the corresponding classification ranges suggested for the characterization of mesotrophic states.[15] If we consider the present exogenous loading conditions, we infer that the exceedance frequency of 15 TP $\mu$g/L is currently greater than 30−35% during the summer period in nearly all segments modeled (Figure 4 and Figure SI-7). In particular, the mean TP levels in Cook's Bay are higher than 20 $\mu$g/L and the corresponding frequency of violations of 15 TP $\mu$g/L is greater than 50%. Similarly, the average chlorophyll *a* concentrations exceed the level of 5 $\mu$g chl*a*/L in the inner parts of both Cook's Bay and Kempenfelt Bay, and the value of 4 $\mu$g chl*a*/L is exceeded more than 25% of the time in the southwestern part of the lake. Our Bayesian network predicts that a 15% phosphorus loading reduction will still result in >25% violations in the inner segments of the two embayments, while the rest of the lake will experience <20% violations. Interestingly, the exceedance

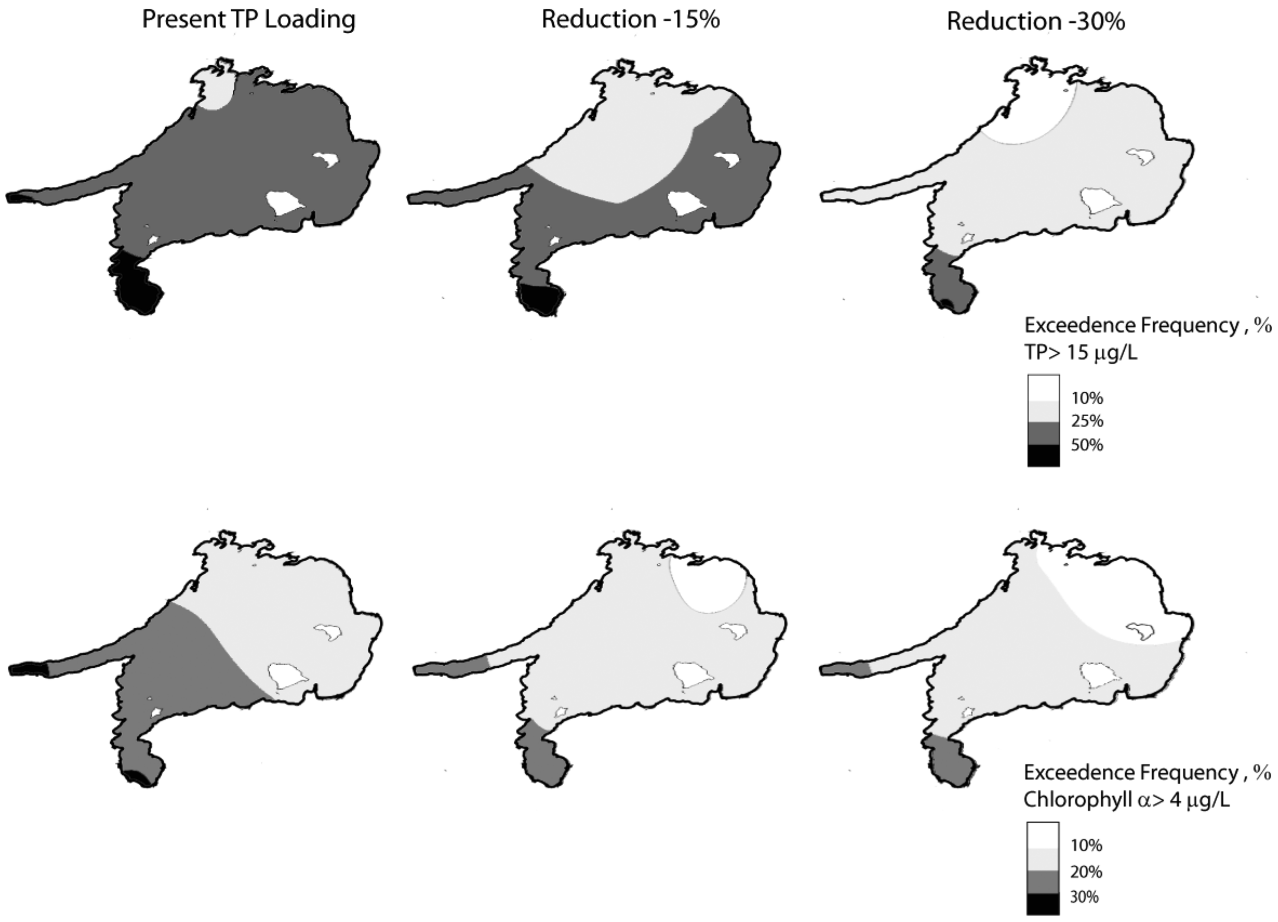



**Figure 4.** Predictions of the Bayesian network for the exceedance frequencies of total phosphorus and chlorophyll *a* thresholds concentrations in Lake Simcoe under the present conditions and two scenarios of phosphorus loading reduction.

frequency of the value of 4 $\mu$g chl*a*/L remained practically unaltered under the scenario of 30% phosphorus loading reduction. The TP levels will decrease in response to the exogenous loading reductions and this improvement will be primarily manifested in the northcentral segments of the system. Our integrated model also predicts that the mean TP values will be lower than 20 $\mu$g/L in Cook's Bay, but the exceedance frequency of the TP level of 15 $\mu$g/L will remain >25%, even if the loading is reduced by 30%.

## 4. DISCUSSION

A major challenge when developing modeling tools aiming to accommodate ecosystem complexity is the ability to combine quantitative descriptions of ecological processes at multiple scales and in a variety of forms (intermediate complexity mathematical models, empirical equations, expert judgments).[21] Bayesian networks offer an effective means for integrating various models into one coherent framework, while rigorously assessing how uncertainties in each component translate to uncertainty in the final predictions. Recent research has shown that this modeling strategy can effectively alleviate problems of spatiotemporal resolution mismatch among different submodels of integrated environmental modeling systems; overcome the conceptual or scale misalignment between processes of interest and supporting information; exploit disparate sources of data that differ with regards to the measurement error and resolution; and accommodate tightly intertwined environmental processes operating at different spatiotemporal scales.[21] Here, we present two models that translate the processes underlying planktonic patterns into an articulated sequence of conditional relationships (TP loading→epilimnetic TP→phytoplankton biomass). This work is relevant to policy and management decisions because it attempts to quantitatively predict ecological changes in Lake Simcoe in response to various phosphorus reduction scenarios.

Young et al. recently re-examined the capacity of an empirical equation historically developed for Lake Simcoe to predict TP concentrations from TP loading rates in the postdreissenid period.[20] Counter to the contemporary paradigm on the impact of dreissenid mussels, the successful application of the model in predicting average, ice-free, whole-lake phosphorus concentrations was interpreted as evidence that their invasion has not fundamentally altered the existing relationship between TP loading and TP concentrations. In this study, our hierarchical spatially explicit model forced with idealized sinusoidal loading relaxes two features of the existing empirical modeling work in Lake Simcoe: (i) the system is no longer assumed thoroughly mixed with uniform concentrations throughout; and (ii) rather than drawing inference upon seasonal average TP values, the ambient TP levels are predicted at finer (monthly, daily) time scales that may be more relevant for addressing the current management issues in Lake Simcoe. In this regard, the plausibility of our estimates of the TP loss rates and the projected differences between the two embayments and the main basin are critical for evaluating the validity of the parametrization obtained. Hiriart-Baer et al. recently reported post-1970 estimates of the gross TP accumulation rates in Cook's Bay and Kempenfelt Bay sediments

at the level of 250−750 mg P/m$^2$/yr and 300 mg P/m$^2$/yr, respectively.[23] Although Hiriart-Baer et al.'s gross TP sedimentation rates are proxies that may not be suitable for deterministic comparisons, our corresponding areal weighted average estimates of the net areal TP loss rates were approximately 300 mg P/m$^2$/yr (C1, C6, C9) and 150 mg P/m$^2$/yr (K39, K42). Further, the likelihood of the unidirectional flow patterns postulated by our model to misrepresent the dilution effects of the water masses from the outer lake suggests that our estimates probably reflect the upper levels of net sedimentation in the two embayments (see Supporting Information Section D). One possible implication of the 2-fold difference between gross and net TP sedimentation rates is that several ecological mechanisms, such as macrophyte growth, transport to the hypolimnion, sediment resuspension and P release, can significantly modulate TP dynamics in the two embayments. Another plausible explanation for the substantial discrepancy between the two studies may simply be the fact that our work spans a shorter time period with improved water quality and significantly lower TP concentrations.[24]

A second regulatory factor of the segment-specific TP budgets is related to our posterior estimates of the flushing times in the bays, which in turn may influence the quantification of the net TP export into the main basin. The hydraulic retention times of the bays can vary significantly on a seasonal basis and are regulated by a suite of factors, such as the wind-induced momentum, the spring freshet, thermocline development, and water temperature gradients.[24] Hydrodynamic modeling calculations based on acoustic Doppler current profiles indicated short retention times between 1 and 2 weeks in the two embayments,[25] or flushing rates that correspond to an equivalent of $\kappa_1 \approx 0.06-0.07$ days$^{-1}$. Although the same values formed the basis for the year-specific $\kappa_1$ priors used in this analysis (eq 3), the derived posterior estimates (0.009−0.012 days$^{-1}$) differed significantly suggesting that they indeed represent the annual net exchange of water in terms of nutrient loadings. Importantly, the relative mismatch between the reported empirical TP accumulation rates of 80−300 mg P/m$^2$/yr in the deeper segments of Lake Simcoe[23] and our predicted net epilimnetic TP loss of 45−85 mg P/m$^2$/yr was fairly similar, and thus the hydrodynamic forcing of the model and consequently our TP mass balance budgets do not seem to introduce systematic bias among the different segments.

Our SEM reinforces earlier empirical modeling work that showed TP is a reliable predictor of the phytoplankton biomass in Lake Simcoe,[20,26] with the robustness of the chlorophyll *a*−TP relationship during the postdreissenid period being verified at nearly all sites of the lake.[20] The only exception was at C1, the shallowest segment in Cook's Bay, where the historical empirical equation overpredicted the standing algal biomass possibly due to increased dreissenid filtering activity. Our SEM results appear to differ somewhat in that this relationship was fairly strong in the C1 segment (0.30 ± 0.11), but weak and poorly identified (C6: −0.02 ± 0.11) in the middle part of Cook's Bay. Yet, this discrepancy should be interpreted with caution as the inclusion of additional covariates in our model along with the use of contemporaneous measurements from individual samplings for all the water quality variables (i.e., no temporal-averaging or lagged relationships were considered) largely determines the relative magnitudes and signs (e.g., negative DIN-phytoplankton) of the various ecological paths derived.[17,18] In Cook's Bay, TP concentrations generally decline moving northward from C1, suggesting that a significant fraction of the TP load entering Cook's Bay from the Holland River is retained in sediment, which was predicted by our model. However, the interesting finding of the present analysis is the substantial interannual variability of the corresponding estimates in segment C6 (Figure SI-4), along with the weak (i.e., highly uncertain) causal links associated with bottom-up (TP→chl*a*) and top-down (herbivorous grazing→chl*a*) factors (Figure 3). The uncertainty characterizing both our models in the middle area of Cook's Bay may be indicative of a dynamic/intermittent environment that could conceivably mask important cause−effect relationships when analyzing daily snapshots from the system. Aside from all the structural and functional changes occurring in this shallow embayment over the last two decades,[9] it is plausible that this uncertainty may also stem from sediment resuspension events, triggered by wind forcing and episodic runoff events, which in turn can have profound influence on local and lakewide biogeochemistry and trophic functioning.[27]

Our analysis highlights the lakewide strength of the relationship between Secchi disk depth values and chlorophyll *a* concentrations. However, it should be noted that while this causal link was originally designed to account for the positive effects of the illumination of the water column on phytoplankton growth (i.e., greater Secchi disk depth suggests adequate light availability, which in turn stimulates algal growth), its negative nature is indicative of an opposite direction path in a strict causal sense, i.e., chlorophyll *a*→Secchi disk depth. The nature of the water clarity−phytoplankton relationship during the period modeled (1999−2007) could have been partly shaped by the colonization of dreissenid mussels.[22] Dreissenids can increase water clarity by reducing algae and other suspended solids through their filtering activity, as well as by reducing whiting events.[12] The improved illumination of the water column has given rise to a dramatic proliferation of submerged aquatic vegetation in the shallow Cook's Bay, and the mean macrophyte biomass along with the areal coverage of submerged aquatic vegetation have increased by approximately 260% and 65%, respectively.[19] The increases in macrophyte biomass and abundance may in turn have induced changes in the local phytoplankton community through a complex array of direct and indirect effects (e.g., enhanced grazing pressure from pelagic zooplankton, changes in nutrient cycling, increased sedimentation, shading, and allelopathy), with a notable example the increase in the abundance of the small flagellate *Cryptomonas*.[26]

Another factor that can affect phytoplankton community structure is the control exerted by herbivorous grazing, and our SEM analysis pinpoints the importance of this causal path throughout the lake (Figure 3). Triggered by the decline of planktivorous fish (smelt and lake herring), the density of large *Daphnia* species (e.g., *Daphnia longiremis*) increased in Lake Simcoe in the late 1980s and may have temporarily contributed to the decrease of phytoplankton biomass.[20] Invasion by the zooplanktivore *Bythotrephes longimanus* may also be indirectly affecting phytoplankton abundance as *Bythotrephes* is known to significantly decrease cladoceran species richness and alter zooplankton behavioral patterns.[26] Our analysis clearly illustrates that future work on the impact of *Bythotrephes* and our understanding of plankton dynamics in Lake Simcoe should be sought in the context of a combined bottom-up and top-down forcing. While not explicitly considered in our analysis, climate warming in the region has been suggested to be responsible for recent trends in the Lake Simcoe thermal structure, such as an increase in the thermal stability of the water column during the ice-free season, an earlier onset of thermal stratification, and delayed fall overturn.[28] The impacts of these changes on lake

chemistry and biota are currently being investigated and there is evidence that the multitude of stressors (invasive species, climate change) along with the contemporary changes in the nutrient loading regime have induced discernible changes in the phytoplankton biomass and community composition. Namely, the chlorophytes and cyanobacteria abundance decreased throughout the lake, while the diatom community composition patterns suggest shifts toward species (e.g., *Fragilaria crotonensis*) that may gain competitive advantage in conditions of increased water column stability and reduced mixing.[26]

One of the benefits of the probabilistic assessment of Lake Simcoe water quality conditions is the ability to optimize water quality monitoring programs and to identify specific areas that require further research efforts. For example, our analysis delineated an area in Cook's Bay where the interplay among exogenous nutrient inputs, hydrodynamics, autotrophic and benthic communities obfuscates the signals of critical relationships between water quality variables of management interest. Our inability to detect significant causal linkages is also reflected in the high model uncertainty in these locations, which in turn explains the predicted moderate response of Cook's Bay to the nutrient loading reductions examined. In this regard, we believe that this finding highlights a critical future research question involving the capacity of wind-driven physical forcing to produce localized "hotspots" of biological productivity in near-shore regions[27] that could conceivably shape the broader impact of dreissenids in Lake Simcoe.[9]

We conclude by reiterating that the present modeling analysis was based on two simple models, which were used because of their ability to remain within the bounds of empirical parameter estimation and therefore accommodate error analysis. In general, however, model-based environmental management is preferred to have stronger mechanistic foundation, as this provides additional assurance that the model will reflect the functional changes in the lake ecosystem induced by the nutrient loading reductions. Yet, adopting a process-based model and invoking extra complexity raises critical questions in regards to the existence of commensurate knowledge of the multifaceted aspects of the Lake Simcoe dynamics or even the capacity to depict them mathematically. Until we can give definitive positive answers to these questions, we believe that the gradual incorporation of complexity, whenever possible and relevant, is the most prudent strategy. Any such model development should be tightly coupled with rigorous assessment of the underlying uncertainty and the Bayesian inference can be an invaluable tool for this purpose.[8]

## ■ ASSOCIATED CONTENT

### Supporting Information

Additional information related to the data set used, the configuration of the models, and the results of our study. This material is available free of charge via the Internet at http://pubs.acs.org.

## ■ AUTHOR INFORMATION

### Corresponding Author

*E-mail: georgea@utsc.utoronto.ca; tel.: +1 416 208 4858; fax: +1 416 287 7279.

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

This project was undertaken with the financial support of the Government of Canada provided through the Department of the Environment. A.G. has also received support from an Ontario Graduate Doctoral Scholarship. All the model codes pertinent to this analysis are available upon request from the corresponding author.

## ■ REFERENCES

(1) Minns, C. K.; Kelso, J. R. M. NO! It is time for a Great Lakes Ecosystem Management Agreement that SUBSUMES the Great Lakes Water Quality Agreement. *J. Great Lakes Res.* **2000**, *26* (1), 1−2.

(2) Krantzberg, G. Science must inform Great Lakes policy. *J. Great Lakes Res.* **2004**, *30* (4), 573−574.

(3) Zhang, W. T.; Arhonditsis, G. B. Predicting the Frequency of Water Quality Standard Violations Using Bayesian Calibration of Eutrophication Models. *J. Great Lakes Res.* **2008**, *34* (4), 698−720.

(4) Mills, E. L.; Casselman, J. M.; Dermott, R.; Fitzsimons, J. D.; Gal, G.; Holeck, K. T.; Hoyle, J. A.; Johannsson, O. E.; Lantry, B. F.; Makarewicz, J. C.; Millard, E. S.; Munawar, I. F.; Munawar, M.; O'Gorman, R.; Owens, R. W.; Rudstam, L. G.; Schaner, T.; Stewart, T. J. Lake Ontario: Food web dynamics in a changing ecosystem (1970−2000). *Can. J Fish. Aquat. Sci.* **2003**, *60* (4), 471−490.

(5) Leon, L. F.; Smith, R. E. H.; Hipsey, M. R.; Bocaniov, S. A.; Higgins, S. N.; Hecky, R. E.; Antenucci, J. P.; Imberger, J. A.; Guildford, S. J. Application of a 3D hydrodynamic-biological model for seasonal and spatial dynamics of water quality and phytoplankton in Lake Erie. *J. Great Lakes Res.* **2011**, *37* (1), 41−53.

(6) Arhonditsis, G. B.; Brett, M. T. Evaluation of the current state of mechanistic aquatic biogeochemical modeling. *Mar. Ecol.: Prog. Ser.* **2004**, *271*, 13−26.

(7) Anderson, T. R. Plankton functional type modelling: Running before we can walk? *J. Plankton Res.* **2005**, *27* (11), 1073−1081.

(8) Arhonditsis, G. B.; Qian, S. S.; Stow, C. A.; Lamon, E. C.; Reckhow, K. H. Eutrophication risk assessment using Bayesian calibration of process-based models: Application to a mesotrophic lake. *Ecol. Model.* **2007**, *208* (2−4), 215−229.

(9) Palmer, M. E.; Winter, J. G.; Young, J. D.; Dillon, P. J.; Guildford, S. J. Introduction and summary of research on Lake Simcoe: Research, monitoring, and restoration of a large lake and its watershed. *J. Great Lakes Res.* **2011**, *37*, 1−6.

(10) Ontario Ministry of the Environment and Lake Simcoe Region Conservation Authority. *Report on the Phosphorus Loads to Lake Simcoe*; OMOE and LSRCA Report, 2009; 18 pp.

(11) Nicholls, K. H. A limnological basis for a Lake Simcoe phosphorus loading objective. *Lake Reserv. Manage.* **1997**, *13* (3), 189−198.

(12) Evans, D. O. Effects of hypoxia on scope-for-activity and power capacity of lake trout (*Salvelinus namaycush*). *Can. J Fish. Aquat. Sci.* **2007**, *64* (2), 345−361.

(13) O'Connor, E. M.; McConnell, C.; Lembcke, D.; Winter, J. G. Estimation of total phosphorus loads for a large, flashy river of a highly developed watershed—seasonal and hysteresis effects. *J. Great Lakes Res.* **2011**, *37*, 26−35.

(14) Cheng, V.; Arhonditsis, G. B.; Brett, M. T. A revaluation of lake-phosphorus loading models using a Bayesian hierarchical framework. *Ecol. Res.* **2010**, *25* (1), 59−76.

(15) ChapraS. C. *Surface Water-Quality Monitoring*; McGraw-Hill Series in Water Resources and Environmental Engineering, 1997.

(16) Arhonditsis, G. B.; Stow, C. A.; Steinberg, L. J.; Kenney, M. A.; Lathrop, R. C.; McBride, S. J.; Reckhow, K. H. Exploring ecological patterns with structural equation modeling and Bayesian analysis. *Ecol. Model.* **2006**, *192* (3−4), 385−409.

(17) Arhonditsis, G. B.; Stow, C. A.; Paerl, H. W.; Valdes-Weaver, L. M.; Steinberg, L. J.; Reckhow, K. H. Delineation of the role of nutrient dynamics and hydrologic forcing on phytoplankton patterns along a freshwater-marine continuum. *Ecol. Model.* **2007**, *208* (2−4), 230−246.

(18) Arhonditsis, G. B.; Paerl, H. W.; Valdes-Weaver, L. M.; Stow, C. A.; Steinberg, L. J.; Reckhow, K. H. Application of Bayesian structural

equation modeling for examining phytoplankton dynamics in the Neuse River Estuary (North Carolina, USA). *Estuarine Coast. Shelf Sci.* **2007**, *72* (1−2), 63−80.
(19) Ginn, B. K. Distribution and limnological drivers of submerged aquatic plant communities in Lake Simcoe (Ontario, Canada): Utility of macrophytes as bioindicators of lake trophic status. *J. Great Lakes Res.* **2011**, *37*, 83−89.
(20) Young, J. D.; Winter, J. G.; Molot, L. A re-evaluation of the empirical relationships connecting dissolved oxygen and phosphorus loading after dreissenid mussel invasion in Lake Simcoe. *J. Great Lakes Res.* **2011**, *37*, 7−14.
(21) Borsuk, M. E.; Stow, C. A.; Reckhow, K. H. A Bayesian network of eutrophication models for synthesis, prediction, and uncertainty analysis. *Ecol. Model.* **2004**, *173* (2−3), 219−239.
(22) Hecky, R. E.; Smith, R. E. H.; Barton, D. R.; Guildford, S. J.; Taylor, W. D.; Charlton, M. N.; Howell, T. The nearshore phosphorus shunt: A consequence of ecosystem engineering by dreissenids in the Laurentian Great Lakes. *Can. J Fish. Aquat. Sci.* **2004**, *61* (7), 1285−1293.
(23) Hiriart-Baer, V. P.; Milne, J. E.; Marvin, C. H. Temporal trends in phosphorus and lacustrine productivity in Lake Simcoe inferred from lake sediment. *J. Great Lakes Res.* **2011**, *37* (4), 764−771.
(24) Eimers, C. M.; Winter, J. G.; Scheider, W. A.; Watmough, S. A.; Nicholls, K. H. Recent changes and patterns in the water chemistry of Lake Simcoe. *J. Great Lakes Res.* **2005**, *31*, 322−332.
(25) W.F. Baird & Associates Coastal Engineers LTD. Oakville, Ontario; *Lake Simcoe Hydrodynamic and Water Quality Model Report*; 2006.
(26) Winter, J. G.; Young, J. D.; Landre, A.; Stainsby, E.; Jarjanazi, H. Changes in phytoplankton community composition of Lake Simcoe from 1980 to 2007 and relationships with multiple stressors. *J. Great Lakes Res.* **2011**, *37*, 63−71.
(27) Johengen, T. H.; Biddanda, B. A.; Cotner, J. B. Stimulation of Lake Michigan plankton metabolism by sediment resuspension and river runoff. *J. Great Lakes Res.* **2008**, *34*, 213−227.
(28) Stainsby, E. A.; Winter, J. G.; Jarjanazi, H.; Paterson, A. M.; Evans, D. O.; Young, J. D. Changes in the thermal stability of Lake Simcoe from 1980 to 2008. *J. Great Lakes Res.* **2011**, *37*, 55−62.

# A CONTINUOUS BAYESIAN NETWORK FOR STUDYING THE CAUSAL LINKS BETWEEN PHOSPHORUS LOADING AND PLANKTON PATTERNS IN LAKE SIMCOE, ONTARIO, CANADA

**(Supporting Information)**

**Alexey Gudimov[1], Eavan O'Connor[2], Maria Dittrich[1], Hamdi Jarjanazi[3], Michelle E. Palmer[3], Eleanor Stainsby[3], Jennifer G. Winter[3], Joelle D. Young[3], George B. Arhonditsis[1]***

[1]Ecological Modeling Laboratory,

Department of Physical & Environmental Sciences, University of Toronto,

Toronto, Ontario, Canada, M1C 1A4

[2]Lake Simcoe Region Conservation Authority

Newmarket, Ontario, Canada, L3Y 4X1

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

* Corresponding author

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

## A) DATA SET DESCRIPTION

Exogenous nutrient loading values during the 1999-2004 period were obtained by the Ministry of the Environment of the Province of Ontario (OMOE), while loading estimates from 2004 to 2007 were provided by the Lake Simcoe Regional Conservation Authority (LSRCA). Non-point sources included tributary loads, polders, and air deposition. Point sources included loadings from septics, Water Control Pollution Stations (WCPC), and urban runoff. Phosphorus inputs from precipitation were assumed proportional to segment areas. The air deposition fluxes were based on both monthly (1999-2004) and daily (2004-2007) estimates. Chlorophyll *α*, *TP, TN, $NO_3$,* and *DIN* concentrations were based on samples collected twice a month during the ice-free period[1]. Composite samples from the euphotic zone were collected from nine (9) stations (Fig. 1a): *ATH* (Atherley Narrows segment), *E51* (Eastern Segment), *K45* and *S15* (Central Segment), *K42* and *K39* (Kempenfelt Bay segment), *C1*, *C6* and *C9* (Cook's Bay segment). Missing data were imputed by linear interpolation between adjacent dates. In cases where the monitoring of a particular variable was discontinued, the data were imputed by linear regressions, in which past information from neighbouring stations was used as predictor variable.

Number of animals and average lengths of identified zooplankton species were used to derive biomass estimates with dry weight-length regression models (Table A1):

$$\ln W = \ln a + b \ln \overline{L} \quad \textbf{(1)}$$

where $\ln W$ is the natural logarithm of the dry weight ($\mu$g), and $\ln \overline{L}$ is the mean of the natural log-transformed lengths (mm) of all the individuals of the same species in a particular sample. Zooplankton biomass estimates in the original scale were obtained by back-transformation and correction for the retransformation bias[2]:

$$W = a \overline{L}^{\,b} exp(0.5\, SEE^2) \quad \textbf{(2)}$$

where *SEE* is the standard error of the model. When the standard error was not provided for a specific regression model, the global mean error from all the regression models was used to fill the missing value. In addition, we accounted for the reduction of dry weight from the chemical preservation of organisms[3] by classifying zooplankton length measurements into two size fractions (large>0.25mm, small<0.25mm). The

derived dry weight estimates from equation (2) were then reduced by 37% and 43% for large and small size animals, respectively.

Based on their feeding patterns, identified species were classified into herbivorous and omnivorous/carnivorous zooplankton. The classification "herbivorous" comprised species that feed upon a variety of small particles including bacteria, algae, ciliates, small rotifers and copepod nauplii as large as 100 $\mu m^4$. With this categorization, meiobenthic cladocerans, such as *Disparalona hamata* and *Pseudochydorus globosus*, were specified as herbivores. Major zooplankton taxonomic groups during the study period (1997-2007) are shown in Fig. SI-8. *Daphnia galeata mendotae* had the largest contribution to the total zooplankton biomass. Calanoid and cyclopoid copepodids also consistently appeared in all sampling stations. These three major groups accounted for approximately 50% of the total zooplankton biomass. Other major species included *Eubosmina longispina*, *Eubosmina coregoni*, *Leptodiaptomus minutes, Skistodiaptomus oregonensis,* and *Daphnia longiremis*. Large zooplankton species, such as *Bythotrephes longimanus* and *Leptodora kindtii*, contribute about 3-4% to the total zooplankton biomass in Lake Simcoe, although some large species are not adequately represented in the routine zooplankton collections due to the size of the net used.

.

**Table A1.** Zooplankton biomass estimation based on dry weight-length regression models collected from the literature. Counts *(N)*, average length *($\bar{L}$)*, and associated length ranges of the different species in Lake Simcoe.

| *Species* | *Category* | *N* | *$\bar{L}$* | *Range* | *ln a* | *b* | *SEE* | *Reference* |
|---|---|---|---|---|---|---|---|---|
| **CLADOCERA** | | | | | | | | |
| *Daphnia ambigua* | Herbivorous | 29 | 0.91 | 0.53-1.48 | 1.54 | 2.29 | 0.0312 | [5] |
| *Daphnia parvula* | Herbivorous | 33 | 0.85 | 0.46-1.24 | 1.08 | 2.16 | 0.0312 | [6] |
| *Daphnia pulicaria* | Herbivorous | 1 | 1.04 | N/A | 1.3585 | 3.14 | 0.0312 | [7] |
| *Daphnia retrocurva* | Herbivorous | 915 | 0.97 | 0.46-1.81 | 0.8637 | 3.1262 | 0.0441 | [8] |
| *Daphnia galeata mendotae* | Herbivorous | 12631 | 0.97 | 0.25-4.00 | 1.0797 | 2.7188 | 0.0352 | [9] |
| *Daphnia longiremis* | Herbivorous | 1294 | 0.87 | 0.34-1.48 | 1.6274 | 3.3367 | 0.0136 | [9] |
| *Daphnia* sp. | Herbivorous | 1 | 0.65 | N/A | 1.0797 | 2.7188 | 0.0352 | [9] |
| *Ceriodaphnia* sp. | Herbivorous | 89 | 0.48 | 0.26-0.75 | 2.7286 | 3.337 | 0.0617 | [8] |
| *Simocephalus serrulatus* | Carnivorous/Omnivorous | 3 | 0.83 | 0.76-0.91 | 1.3863 | 3.81 | 0.0312 | [5] |
| *Bosmina freyi, liederi* or *longispina* | Herbivorous | 3102 | 0.36 | 0.16-0.72 | 3.5274 | 3.5859 | 0.0936 | [10] |
| *Bosmina longirostris* | Herbivorous | 6569 | 0.37 | 0.21-1.15 | 2.4751 | 3.3614 | 0.0083 | [8] |
| *Eubosmina coregoni* | Herbivorous | 5089 | 0.47 | 0.24-1.19 | 3.0871 | 2.3371 | 0.0046 | [11] |
| *Eubosmina longispina* | Herbivorous | 7339 | 0.44 | 0.15-0.89 | 3.5274 | 3.5859 | 0.0936 | [10] |
| *Alona affinis* | Herbivorous | 2 | 0.70 | 0.48-0.91 | 2.7676 | 3.84 | 0.0312 | [5] |
| *Alona guttata* | Herbivorous | 1 | 0.34 | N/A | 2.2367 | 2.7418 | 0.0231 | [8] |
| *Alona* sp. | Herbivorous | 12 | 0.48 | 0.29-0.78 | 2.2367 | 2.7418 | 0.0231 | [8] |
| *Disparalona hamata* | Herbivorous | 1 | 0.25 | N/A | 3.5276 | 3.264 | 0.0312 | [12] |
| *Camptocercus* sp. | Herbivorous | 3 | 0.76 | 0.69-0.81 | Log(a)=9.05 | 0.85 | 0.0312 | [5] |
| *Chydorus sphaericus* | Herbivorous | 155 | 0.28 | 0.15-0.51 | 3.127 | 3.3678 | 0.0011 | [8] |
| *Pseudochydorus globosus* | Herbivorous | 53 | 0.53 | 0.37-0.75 | 3.127 | 3.3678 | 0.0312 | [8] |
| *Diaphanosoma birgei* | Herbivorous | 575 | 0.37 | 0.005-1.12 | 1.072 | 2.9054 | 0.0378 | [8] |
| *Diaphanosoma brachyurum* | Herbivorous | 75 | 0.76 | 0.38-1.08 | 1.6242 | 3.0468 | 0.1370 | [10] |
| *Sida crystallina* | Herbivorous | 6 | 0.89 | 0.75-1.08 | 2.0539 | 2.189 | 0.0312 | [12] |
| *Holopedium* sp. | Carnivorous/Omnivorous | 43 | 0.80 | 0.43-1.16 | 2.788 | 3.2102 | 0.0515 | [8] |
| *Polyphemus pediculus* | Carnivorous/Omnivorous | 12 | 0.75 | 0.35-1.13 | 2.7792 | 2.152 | 0.0312 | [12] |
| *Bythotrephes longimanus* | Carnivorous/Omnivorous | 348 | 2.91 | 0.74-5.76 | Log(a)=1.23 | 2.09 | 0.0312 | [13] |
| *Leptodora kindtii* | Carnivorous/Omnivorous | 124 | 4.99 | 0.96-12* | 0.445 | 1.873 | 0.0037 | [11] |

| Species Observed | Diet Category | *N* | $\bar{L}$ | *Range* | *ln a* | *b* | *SEE* | *Reference* |
|---|---|---|---|---|---|---|---|---|
| **COPEPODA** | | | | | | | | |
| Calanoid copepodids | Herbivorous | 22755 | 0.66 | 0.21-1.84 | 0.9799 | 2.7765 | 0.0093 | [8] |
| Calanoid nauplius | Herbivorous | 17321 | 0.20 | 0.06-1.11 | 0.9926 | 2.0997 | 0.0172 | [8] |
| *Leptodiaptomus minutus* | Herbivorous | 13506 | 0.82 | 0.15-1.31 | 1.0377 | 2.8255 | 0.0157 | [8] |
| *Leptodiaptomus sicilis* | Herbivorous | 450 | 1.23 | 0.88-1.52 | 0.9311 | 3.036 | 0.0026 | [8] |
| *Skistodiaptomus oregonensis* | Herbivorous | 4115 | 1.13 | 0.70-1.59 | 0.9717 | 2.7323 | 0.0071 | [8] |
| *Epischura lacustris* | Carnivorous/Omnivorous | 1264 | 1.51 | 0.91-2.02 | 1.2702 | 2.485 | 0.0116 | [8] |
| *Epischura lacustris* (copepodids) | Carnivorous/Omnivorous | 2112 | 0.82 | 0.29-1.61 | 0.8655 | 2.9373 | 0.0086 | [8] |
| Cyclopoid copepodids | Carnivorous/Omnivorous | 30878 | 0.50 | 0.15-1.20 | 1.03 | 2.505 | 0.0479 | [8] |
| Cyclopoid nauplius | Herbivorous | 25848 | 0.16 | 0.06-0.44 | 1.6388 | 2.4474 | 0.0196 | [8] |
| *Acanthocyclops robustus* | Carnivorous/Omnivorous | 12 | 0.74 | 0.63-1.01 | 1.2177 | 3.4934 | 0.0199 | [8] |
| *Acanthocyclops vernalis* | Carnivorous/Omnivorous | 28 | 0.91 | 0.7-1.23 | 1.2177 | 3.4934 | 0.0199 | [8] |
| *Cyclops scutifer* | Carnivorous/Omnivorous | 2 | 0.92 | 0.53-1.31 | 1.2286 | 2.6398 | 0.0782 | [10] |
| *Diacyclops bicuspidatus thomasi* | Carnivorous/Omnivorous | 12108 | 0.80 | 0.52-1.20 | 0.8066 | 4.0823 | 0.0173 | [8] |
| *Eucyclops neomacruroides* | Carnivorous/Omnivorous | 1 | 1.02 | N/A | 1.1615 | 2.9559 | 0.0263 | [8] |
| *Eucyclops serrulatus* | Carnivorous/Omnivorous | 13 | 0.70 | 0.6-0.89 | 1.1615 | 2.9559 | 0.0263 | [8] |
| *Macrocyclops albidus* | Carnivorous/Omnivorous | 1 | 0.65 | N/A | 1.3169 | 2.7917 | 0.0254 | [8] |
| *Mesocyclops americanus* | Carnivorous/Omnivorous | 2 | 1.16 | 1.12-1.19 | 1.3169 | 2.7917 | 0.0254 | [8] |
| *Mesocyclops edax* | Carnivorous/Omnivorous | 4072 | 0.90 | 0.58-1.44 | 1.3169 | 2.7917 | 0.0254 | [8] |
| *Tropocyclops extensus* | Herbivorous | 3003 | 0.49 | 0.35-0.69 | 1.1615 | 2.9559 | 0.0263 | [8] |
| *Tropocyclops prasinus* | Herbivorous | 3735 | 0.48 | 0.35-1.01 | 1.1615 | 2.9559 | 0.0263 | [8] |
| Harpacticoid sp. | Herbivorous | 16 | 0.47 | 0.21**-0.62 | -5.32 | 13.95 | 0.0312 | [14] |

* The maximum length was set equal to 12 mm to correct the measurement error
** The minimum length was set equal to 0.21 mm to avoid negative values

## References


1. Young, J. D.; Winter, J. G.; Molot, L., A re-evaluation of the empirical relationships connecting dissolved oxygen and phosphorus loading after dreissenid mussel invasion in Lake Simcoe. *Journal of Great Lakes Research* **2011,** *37*, 7-14.
2. Stow, C.A.; Reckhow, K.H.; Qian S.S. A Bayesian approach to retransformation bias in transformed regression. *Ecology* **2006,** *87*: 1472-1477.
3. Giguere, L. A.; St-Pierre, J. F.; Bernier, B.; Vezina, A.; Rondeau, J. G., Can we estimate the true weight of zooplankton samples after chemical preservation? *Canadian Journal of Fisheries and Aquatic Sciences* **1989,** *46*, 522-527.
4. Porter, K. G.; Gerritsen, J.; Orcutt, J. D., The effect of food concentration on swimming patterns, feeding-behavior, ingestion, assimilation, and respiration by Daphnia. *Limnology & Oceanography* **1982,** *27*, 935-949.
5. Dumont, H. J.; Vandevelde, I.; Dumont, S. Dry weight estimate of biomass in a selection of cladocera, copepoda and rotifera from plankton, periphyton and benthos of continental waters. *Oecologia* **1975,** *19*, 75-97.
6. Pace, M. L.; Orcutt, J. D., The relative importance of protozoans, rotifers, and crustaceans in a fresh-water zooplankton community. *Limnology & Oceanography* **1981,** *26*, 822-830.
7. Snow, N. B., Effect of season and animal size on caloric content of daphnia-pulicaria forbes. *Limnology & Oceanography* **1972,** *17*, 909-913.
8. Malley, D. F.; Lawrence, S. G.; Maciver, M. A.; Findlay, W. J., Range of variation in estimates of dry weight for planktonic crustacea and rotifera from temperate North American lakes. *Canadian Technical Report of Fisheries and Aquatic Sciences* **1989,** (1666), I.
9. Lawrence, S. G.; Malley, D. F.; Findlay, W. J.; Maclver, M. A.; Delbaere, I. L., Method for estimating dry-weight of fresh-water planktonic crustaceans from measures of length and shape. *Canadian Journal of Fisheries and Aquatic Sciences* **1987,** *44*, 264-274.
10. Bottrell, H. H.; Duncan, A.; Gliwicz, Z. M.; Grygierek, E.; Herzig, A.; Hillbricht-Ilkowska, A.; Kurasawa, H.; Larsson, P.; Weglenska, T., Review of some problems in zooplankton production studies. *Norwegian Journal of Zoology* **1976,** *24*, 419-456.
11. Culver, D. A.; Boucherle, M. M.; Bean, D. J.; Fletcher, J. W., Biomass of fresh-water crustacean zooplankton from length weight regressions. *Canadian Journal of Fisheries and Aquatic Sciences* **1985,** *42*, 1380-1390.
12. Rosen, R. A., Length-dry weight relationships of some freshwater zooplankton. *Journal of Freshwater Ecology* **1981,** *1*, 225-229.
13. Makarewicz, J. C.; Jones, H. D., Occurrence of bythotrephes-cederstroemi in lake-ontario offshore waters. *Journal of Great Lakes Research* **1990,** *16*, 143-147.
14. Goodman, K. S., The estimation of individual dry-weight and standing crop of harpacticoid copepods. *Hydrobiologia* **1980,** *72*, 253-259.

**B) FIGURES LEGENDS**

**Figure SI-1:** Conceptual model of the epilimnetic plankton dynamics in Lake Simcoe.

**Figure SI-2:** Comparison of the observed and mean predicted (along with 95% credible intervals) total phosphorus concentrations in the nine segments of the continuously-stirred-talk reactor (CSTR) model, forced with idealized sinusoidal loading.

**Figure SI-3:** Scatter plots of the segment-specific sedimentation rates against the residence times for the year 2007.

**Figure SI-4:** Posterior estimates of the total phosphorus sedimentation rates ($year^{-1}$) at the different segments of Lake Simcoe.

**Figure SI-5:** Predicted annual phosphorus mass balance and net exchange rates at the different segments of Lake Simcoe during the study period (1999-2007).

**Figure SI-6:** Comparison of the observed and mean predicted (along with 95% credible intervals) chlorophyll *a* and zooplankton biomass values in the nine segments of the Lake Simcoe structural equation model.

**Figure SI-7:** Predictions of the Bayesian network for the average total phosphorus and chlorophyll *a* concentrations in Lake Simcoe under the present conditions and two scenarios of phosphorus loading reduction.

**Figure SI-8**: Zooplankton composition by biomass in Lake Simcoe during the study period (1997-2007).

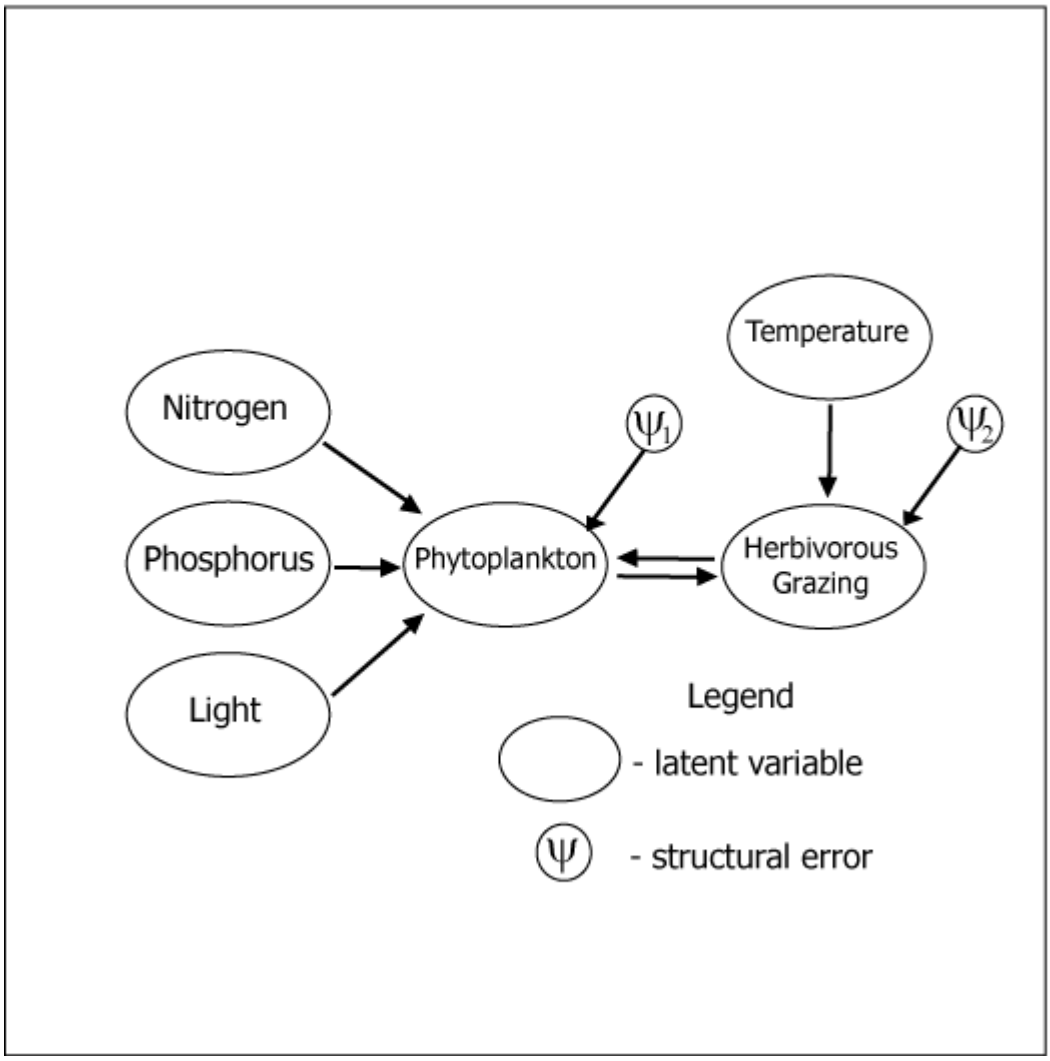

Temperature
Nitrogen
ψ1
ψ2
Phosphorus
Phytoplankton
Herbivorous
Grazing
Light
Legend
- latent variable
ψ
- structural error

**Figure SI-1**

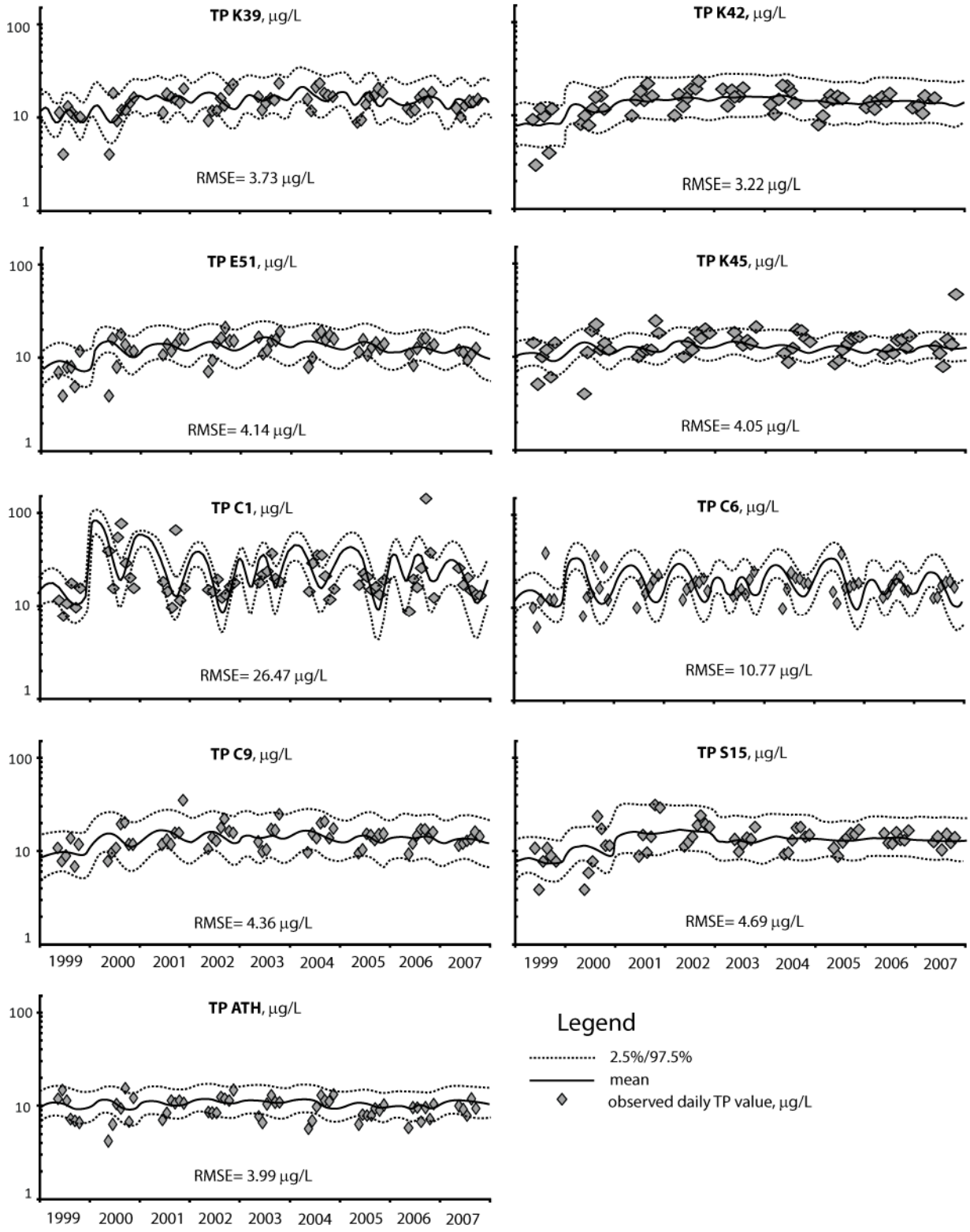

TP K39, µg/L
RMSE= 3.73 µg/L
TP K42, µg/L
RMSE= 3.22 µg/L
TP E51, µg/L
RMSE= 4.14 µg/L
TP K45, µg/L
RMSE= 4.05 µg/L
TP C1, µg/L
RMSE= 26.47 µg/L
TP C6, µg/L
RMSE= 10.77 µg/L
TP C9, µg/L
RMSE= 4.36 µg/L
TP S15, µg/L
RMSE= 4.69 µg/L
TP ATH, µg/L
RMSE= 3.99 µg/L
100
10
1
1999 2000 2001 2002 2003 2004 2005 2006 2007
Legend
2.5%/97.5%
mean
observed daily TP value, µg/L

**Figure SI-2**

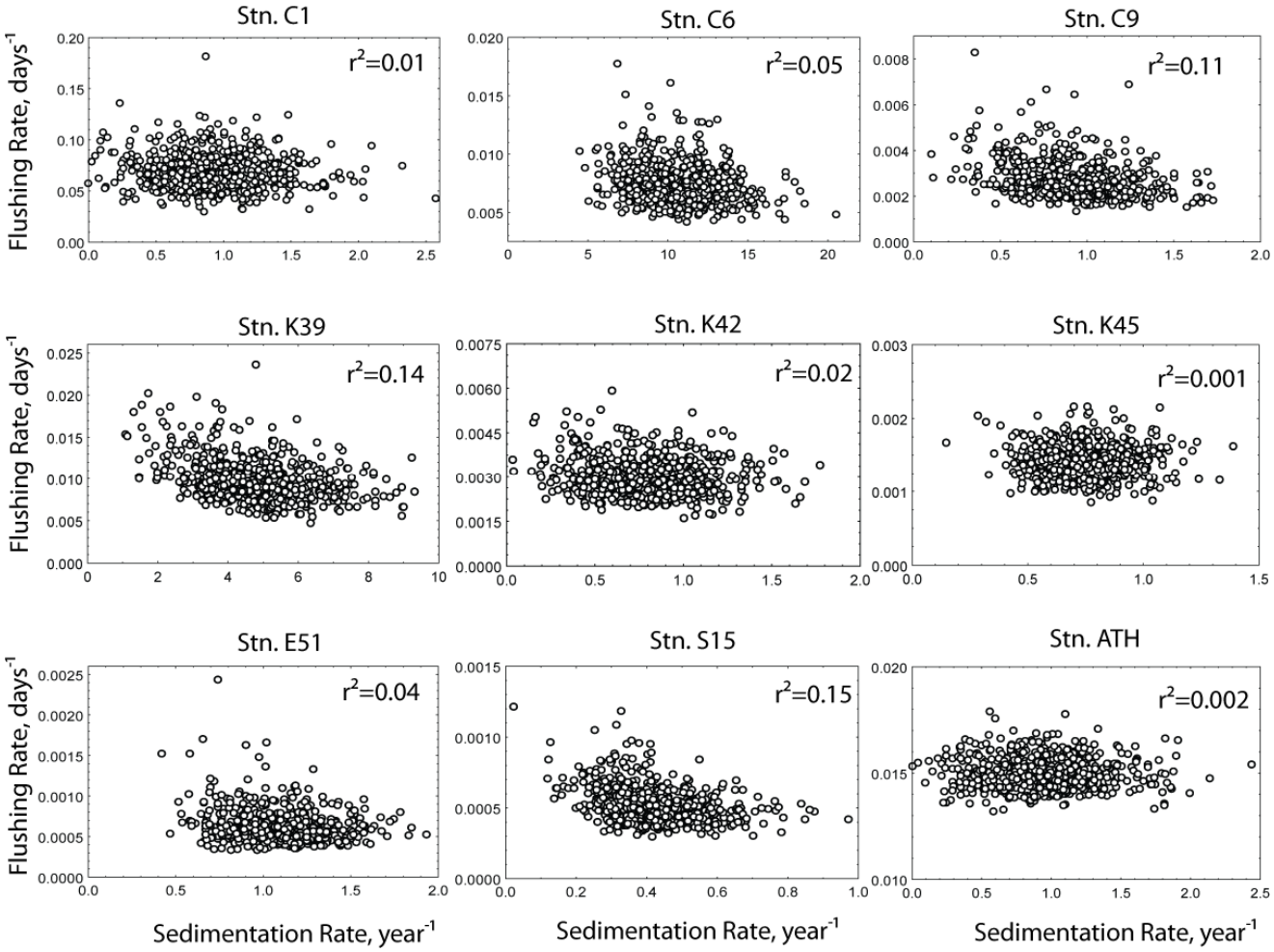

Stn. C1
r²=0.01
Stn. C6
r²=0.05
Stn. C9
r²=0.11
Stn. K39
r²=0.14
Stn. K42
r²=0.02
Stn. K45
r²=0.001
Stn. E51
r²=0.04
Stn. S15
r²=0.15
Stn. ATH
r²=0.002
Flushing Rate, days⁻¹
Sedimentation Rate, year⁻¹

**Figure SI-3**

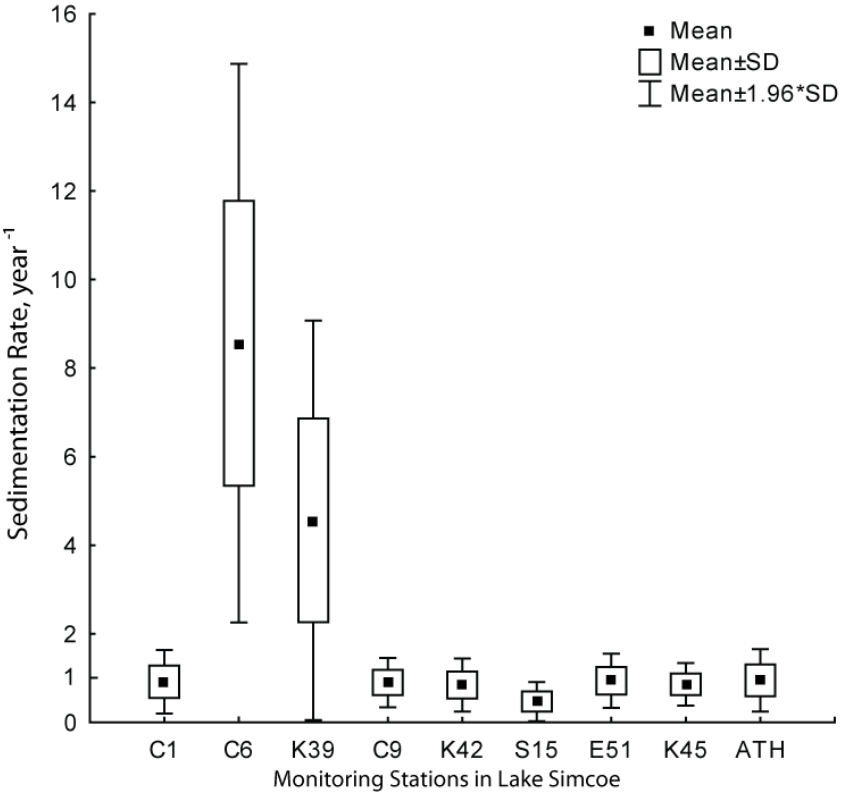

Mean
Mean±SD
Mean±1.96*SD
Sedimentation Rate, year -1
16
14
12
10
8
6
4
2
1
0
C1
C6
K39
C9
K42
S15
E51
K45
ATH
Monitoring Stations in Lake Simcoe

**Figure SI-4**

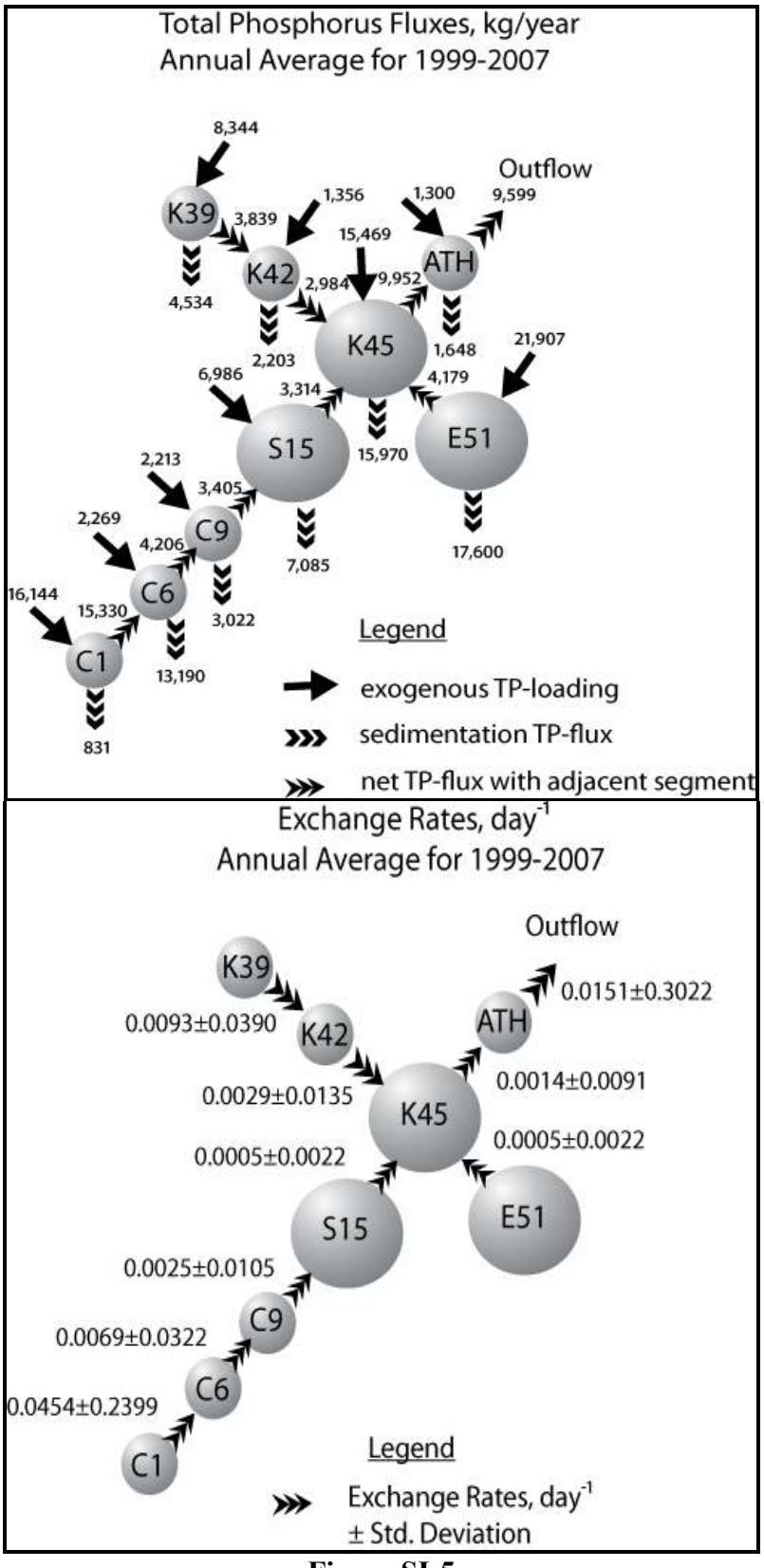

Total Phosphorus Fluxes, kg/year
Annual Average for 1999-2007
8,344
Outflow
K39
1,356
1,300
9,599
3,839
15,469
ATH
K42
2,984
9,952
4,534
2,203
K45
1,648
21,907
6,986
4,179
3,314
S15
E51
15,970
2,213
3,405
C9
2,269
4,206
7,085
17,600
C6
16,144
15,330
3,022
Legend
C1
13,190
exogenous TP-loading
831
sedimentation TP-flux
net TP-flux with adjacent segment
Exchange Rates, day-1
Annual Average for 1999-2007
Outflow
K39
0.0151±0.3022
0.0093±0.0390
K42
ATH
0.0014±0.0091
0.0029±0.0135
K45
0.0005±0.0022
0.0005±0.0022
S15
E51
0.0025±0.0105
C9
0.0069±0.0322
C6
0.0454±0.2399
Legend
C1
Exchange Rates, day-1
± Std. Deviation

**Figure SI-5**

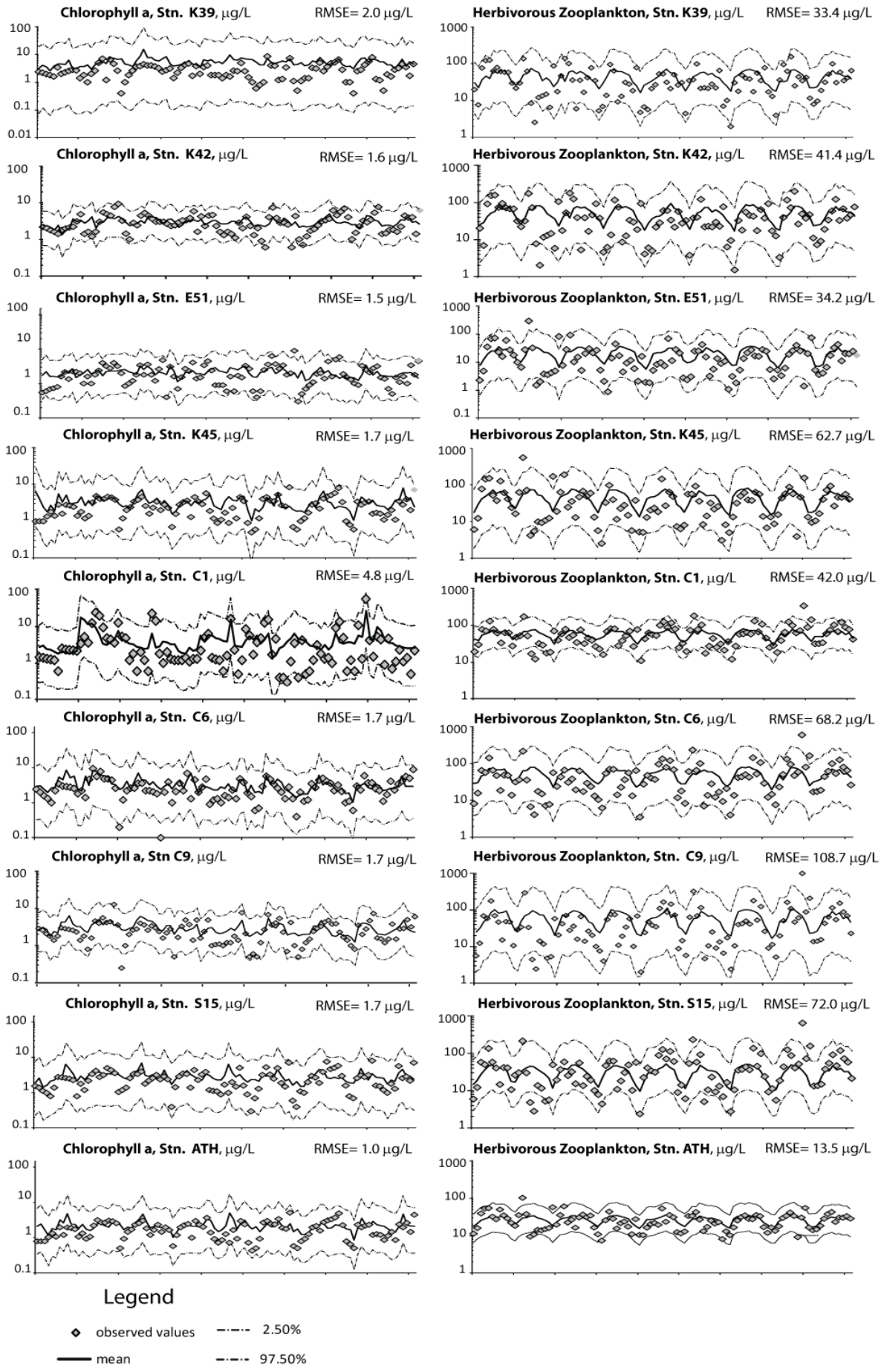

Chlorophyll a, Stn. K39, µg/L
RMSE= 2.0 µg/L
Herbivorous Zooplankton, Stn. K39, µg/L
RMSE= 33.4 µg/L
Chlorophyll a, Stn. K42, µg/L
RMSE= 1.6 µg/L
Herbivorous Zooplankton, Stn. K42, µg/L
RMSE= 41.4 µg/L
Chlorophyll a, Stn. E51, µg/L
RMSE= 1.5 µg/L
Herbivorous Zooplankton, Stn. E51, µg/L
RMSE= 34.2 µg/L
Chlorophyll a, Stn. K45, µg/L
RMSE= 1.7 µg/L
Herbivorous Zooplankton, Stn. K45, µg/L
RMSE= 62.7 µg/L
Chlorophyll a, Stn. C1, µg/L
RMSE= 4.8 µg/L
Herbivorous Zooplankton, Stn. C1, µg/L
RMSE= 42.0 µg/L
Chlorophyll a, Stn. C6, µg/L
RMSE= 1.7 µg/L
Herbivorous Zooplankton, Stn. C6, µg/L
RMSE= 68.2 µg/L
Chlorophyll a, Stn C9, µg/L
RMSE= 1.7 µg/L
Herbivorous Zooplankton, Stn. C9, µg/L
RMSE= 108.7 µg/L
Chlorophyll a, Stn. S15, µg/L
RMSE= 1.7 µg/L
Herbivorous Zooplankton, Stn. S15, µg/L
RMSE= 72.0 µg/L
Chlorophyll a, Stn. ATH, µg/L
RMSE= 1.0 µg/L
Herbivorous Zooplankton, Stn. ATH, µg/L
RMSE= 13.5 µg/L
Legend
observed values
2.50%
mean
97.50%

**Figure SI-6**

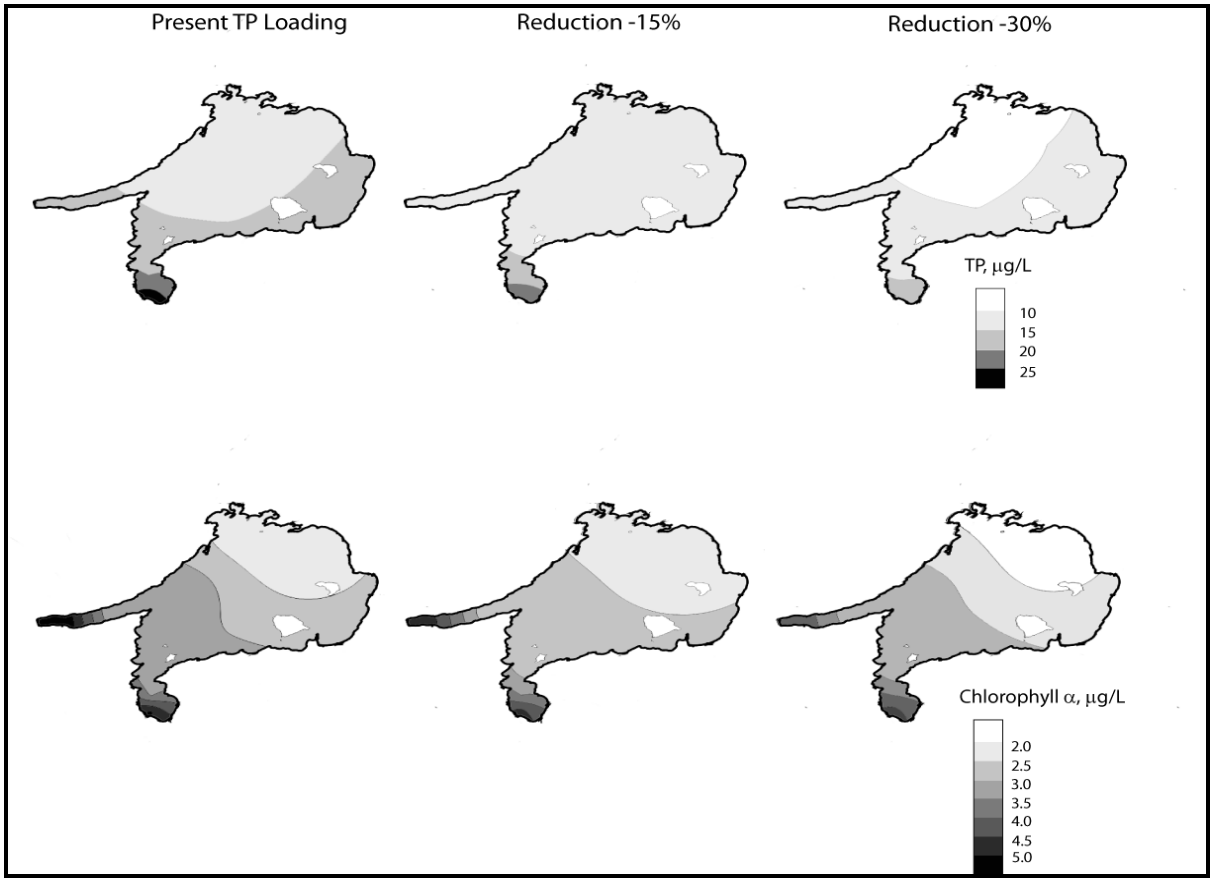

Present TP Loading
Reduction -15%
Reduction -30%
TP, µg/L
10
15
20
25
Chlorophyll α, µg/L
2.0
2.5
3.0
3.5
4.0
4.5
5.0

**Figure SI-7**

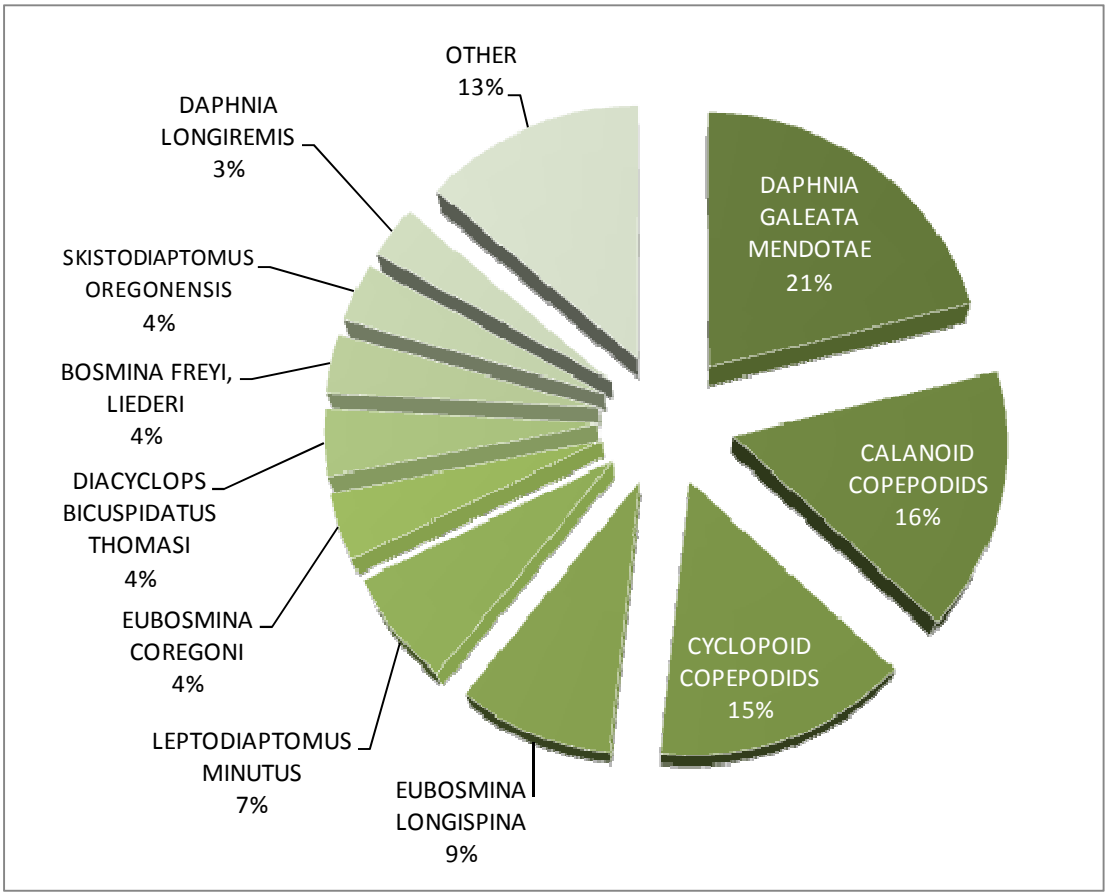

OTHER
13%
DAPHNIA
LONGIREMIS
3%
SKISTODIAPTOMUS
OREGONENSIS
4%
BOSMINA FREYI,
LIEDERI
4%
DIACYCLOPS
BICUSPIDATUS
THOMASI
4%
EUBOSMINA
COREGONI
4%
LEPTODIAPTOMUS
MINUTUS
7%
EUBOSMINA
LONGISPINA
9%
CYCLOPOID
COPEPODIDS
15%
CALANOID
COPEPODIDS
16%
DAPHNIA
GALEATA
MENDOTAE
21%

**Figure SI-8**

## C) TABLES

**Table SI-1:** Continuously Stirred Tank Reactor Model: Summary statistics of the posterior parameter distributions for the nine segments in Lake Simcoe.

| | Cook's Bay | | | | | | Kempenfelt Bay | | | | Main Basin | | | | East End | | Outflow | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | C1 | | C6 | | C9 | | K39 | | K42 | | S15 | | K45 | | E51 | | ATH | |
| Stochastic Nodes | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| $\kappa_{2\ 1999j}$ | 0.91 | 0.37 | 8.42 | 1.78 | 0.92 | 0.27 | 7.40 | 1.68 | 1.04 | 0.30 | 0.77 | 0.19 | 1.03 | 0.20 | 1.39 | 0.25 | 0.89 | 0.35 |
| $\kappa_{2\ 2000j}$ | 0.90 | 0.36 | 8.81 | 1.72 | 0.99 | 0.27 | 6.04 | 1.66 | 0.89 | 0.29 | 0.72 | 0.17 | 1.00 | 0.19 | 1.15 | 0.21 | 0.99 | 0.37 |
| $\kappa_{2\ 2001j}$ | 0.92 | 0.37 | 7.02 | 1.60 | 0.82 | 0.28 | 2.60 | 1.13 | 0.74 | 0.28 | 0.23 | 0.10 | 0.66 | 0.16 | 0.65 | 0.16 | 0.91 | 0.35 |
| $\kappa_{2\ 2002j}$ | 0.91 | 0.37 | 3.67 | 0.98 | 0.79 | 0.27 | 2.84 | 1.16 | 0.73 | 0.29 | 0.21 | 0.09 | 0.60 | 0.15 | 0.63 | 0.15 | 0.91 | 0.35 |
| $\kappa_{2\ 2003j}$ | 0.92 | 0.36 | 8.02 | 1.93 | 0.85 | 0.27 | 2.98 | 1.20 | 0.75 | 0.28 | 0.43 | 0.14 | 0.78 | 0.17 | 0.96 | 0.24 | 0.96 | 0.36 |
| $\kappa_{2\ 2004j}$ | 0.92 | 0.36 | 6.46 | 1.48 | 0.88 | 0.27 | 2.30 | 0.97 | 0.81 | 0.27 | 0.44 | 0.13 | 0.78 | 0.16 | 0.68 | 0.15 | 0.93 | 0.35 |
| $\kappa_{2\ 2005j}$ | 0.93 | 0.37 | 13.14 | 2.29 | 1.03 | 0.29 | 6.30 | 1.56 | 0.93 | 0.29 | 0.53 | 0.14 | 1.12 | 0.19 | 1.03 | 0.19 | 1.02 | 0.37 |
| $\kappa_{2\ 2006j}$ | 0.93 | 0.36 | 10.68 | 2.47 | 0.94 | 0.28 | 6.11 | 1.59 | 0.87 | 0.29 | 0.48 | 0.13 | 1.00 | 0.18 | 0.89 | 0.17 | 1.00 | 0.36 |
| $\kappa_{2\ 2007j}$ | 0.92 | 0.36 | 10.77 | 2.34 | 0.91 | 0.28 | 4.59 | 1.46 | 0.83 | 0.29 | 0.41 | 0.13 | 0.76 | 0.17 | 1.10 | 0.24 | 0.94 | 0.36 |
| $\sigma_{\kappa 2j}$ | 0.36 | 0.09 | 6.31 | 1.33 | 0.32 | 0.07 | 3.29 | 0.79 | 0.33 | 0.08 | 0.42 | 0.09 | 0.31 | 0.06 | 0.34 | 0.08 | 0.36 | 0.10 |
| $\tau$ | 0.39, 0.01 | | | | | | | | | | | | | | | | | |
| $\kappa_2$ | 0.86, 0.07 | | | | | | | | | | | | | | | | | |
| $\kappa_{2\mu}$ | 0.78, 0.27 | | | | | | | | | | | | | | | | | |
| $\sigma_{\kappa 2}$ | 0.36, 0.10 | | | | | | | | | | | | | | | | | |

**Table SI-2:** Structural Equation Model: Summary statistics of the posterior parameter distributions for the nine segments in Lake Simcoe.

| | C1 | | C6 | | C9 | | K39 | | K42 | | S15 | | K45 | | E51 | | ATH | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Parameters | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| $\gamma_{(Nitrogen)j}$ | 0.13 | 0.16 | -0.22 | 0.15 | -0.36 | 0.15 | -0.08 | 0.18 | -0.36 | 0.16 | 0.00 | 0.14 | -0.34 | 0.17 | -0.20 | 0.12 | 0.08 | 0.14 |
| $\gamma_{(Phosphorus)j}$ | 0.30 | 0.11 | -0.02 | 0.11 | 0.27 | 0.09 | 0.27 | 0.11 | 0.29 | 0.10 | 0.17 | 0.10 | 0.20 | 0.10 | 0.46 | 0.10 | 0.10 | 0.10 |
| $\gamma_{(Light)j}$ | -0.42 | 0.11 | -0.38 | 0.11 | -0.38 | 0.11 | -0.18 | 0.13 | -0.26 | 0.11 | -0.37 | 0.13 | -0.28 | 0.11 | -0.32 | 0.10 | -0.47 | 0.12 |
| $\gamma_{(Temperature)j}$ | 0.42 | 0.10 | 0.43 | 0.11 | 0.47 | 0.11 | 0.46 | 0.11 | 0.46 | 0.11 | 0.48 | 0.10 | 0.50 | 0.10 | 0.49 | 0.10 | 0.50 | 0.11 |
| $\beta_{(Zooplankton)j}$ | -0.18 | 0.10 | -0.12 | 0.10 | -0.20 | 0.09 | -0.23 | 0.10 | -0.20 | 0.09 | -0.10 | 0.10 | -0.19 | 0.09 | -0.20 | 0.09 | -0.23 | 0.10 |
| $\beta_{(Chlorophyll)j}$ | 0.06 | 0.04 | 0.04 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.05 | 0.04 | 0.03 | 0.03 | 0.04 | 0.03 | 0.03 | 0.03 |
| $\lambda_{(TN)j}$ | 0.88 | 0.20 | 0.32 | 0.12 | 0.15 | 0.10 | 0.10 | 0.08 | 0.25 | 0.12 | 0.45 | 0.15 | 0.28 | 0.13 | 0.55 | 0.12 | 0.16 | 0.10 |
| $var(\delta_{DIN})_j$ | 0.44 | 0.19 | 0.31 | 0.15 | 0.31 | 0.16 | 0.38 | 0.17 | 0.29 | 0.15 | 0.35 | 0.17 | 0.42 | 0.18 | 0.29 | 0.15 | 0.36 | 0.18 |
| $var(\delta_{TN})_j$ | 0.65 | 0.16 | 0.99 | 0.07 | 1.03 | 0.07 | 1.05 | 0.08 | 0.99 | 0.07 | 0.94 | 0.07 | 0.99 | 0.07 | 0.88 | 0.06 | 0.90 | 0.07 |
| $var(\delta_{TP})_j$ | 0.32 | 0.16 | 0.28 | 0.15 | 0.31 | 0.15 | 0.34 | 0.16 | 0.32 | 0.16 | 0.33 | 0.17 | 0.33 | 0.15 | 0.21 | 0.11 | 0.26 | 0.14 |
| $var(\delta_{Light})_j$ | 0.29 | 0.16 | 0.29 | 0.14 | 0.32 | 0.16 | 0.35 | 0.16 | 0.36 | 0.16 | 0.25 | 0.13 | 0.27 | 0.15 | 0.27 | 0.14 | 0.23 | 0.11 |
| $var(\delta_{Temperature})_j$ | 0.31 | 0.15 | 0.31 | 0.14 | 0.31 | 0.15 | 0.29 | 0.15 | 0.31 | 0.15 | 0.29 | 0.15 | 0.30 | 0.15 | 0.28 | 0.15 | 0.23 | 0.11 |
| $var(\varepsilon_{phytoplankton})_j$ | 0.20 | 0.10 | 0.39 | 0.16 | 0.19 | 0.09 | 0.47 | 0.17 | 0.28 | 0.13 | 0.43 | 0.16 | 0.30 | 0.15 | 0.22 | 0.11 | 0.33 | 0.16 |
| $var(\varepsilon_{zooplankton})_j$ | 0.25 | 0.13 | 0.31 | 0.14 | 0.34 | 0.14 | 0.29 | 0.14 | 0.34 | 0.15 | 0.26 | 0.14 | 0.29 | 0.14 | 0.28 | 0.13 | 0.22 | 0.12 |
| $\varphi(Nitrogen)_j$ | 0.84, 0.08 | | | | | | | | | | | | | | | | | |
| $\varphi(TP)$ | 0.90, 0.06 | | | | | | | | | | | | | | | | | |
| $\varphi(Light)$ | 0.85, 0.06 | | | | | | | | | | | | | | | | | |
| $\varphi(Temp)$ | 0.89, 0.06 | | | | | | | | | | | | | | | | | |
| $\psi_1$ | 0.65, 0.06 | | | | | | | | | | | | | | | | | |
| $\psi_2$ | 0.70, 0.06 | | | | | | | | | | | | | | | | | |

## D) FEEDFORWARD SYSTEM OF CONTINUOUSLY-STIRRED TANK REACTORS WITH SINUSOIDAL LOADING FORCING

**D1) Model assumptions and limitations:** The feedforward series of completely mixed reactors is a pragmatic approach to accommodate the spatial variability in Lake Simcoe, given the lack of hydrodynamic information from the system to support a more sophisticated modeling construct, e.g., lack of chloride concentrations to properly constrain the intersegment mixing processes during the study period. Relative to the feedback framework (i.e., bidirectional mass exchange among the spatial compartments), the feedforward setup is founded upon the assumption of a net unidirectional flow, thereby simplifying both steady-state and time-variable solutions. Yet, while this approach is conceptually suitable to model horizontal mass exchanges in a chain of lakes or a stream, its validity to accommodate the spatial heterogeneity of a single lake holds true only under certain conditions. In particular, the net mass of phosphorus transported from segment A to segment B is:

$$\frac{-Q_A C_A + Q_B C_B}{V_A} = \frac{-Q_{NET}}{V_A} C_A \qquad \textbf{(Eq. D1)}$$

$$\text{or} \qquad Q_{NET} = Q_A - Q_B \frac{C_B}{C_A} > 0 \qquad \textbf{(Eq. D2)}$$

where $Q_A$ is the flow from segment *A* to segment *B*; $Q_B$ is the flow from segment *B* to *A*; $Q_{NET}$ is the net flow exchange among the two segments; *V* represents the volume of a particular segment; and *C* is the *TP* concentration in any segment.

In Lake Simcoe, the feedforward spatial configuration postulates three chains of reactors that connect Cook's Bay (*C1→C6→C9→S15→K45*), Kempenfelt Bay (*K39→K42→K45*), and the eastern segment *(E51→K45*) with the main basin of the lake, which subsequently discharges to Atherley Narrows (*K45→ATH*).

**Table D1:** May-October average total phosphorus concentrations (µg/L), epilimnetic volumes ($10^6$ $m^3$), and outflows ($10^3$ $m^3$/day) of the nine segments during the study period (1999-2007).

| Station | *C1* | *C6* | *C9* | *S15* | *K39* | *K42* | *E51* | *K45* | *ATH* |
|---|---|---|---|---|---|---|---|---|---|
| *TP* | | | | | | | | | |
| Average | 23.4 | 17.2 | 14.5 | 13.7 | 14.2 | 14.1 | 12.9 | 13.7 | 9.9 |
| StDev | 21.1 | 8.0 | 5.0 | 5.2 | 4.4 | 4.5 | 4.4 | 5.5 | 3.0 |
| *Volume* | 29.3 | 81.8 | 238.0 | 1296.0 | 68.8 | 197.1 | 1542.0 | 1533.7 | 160.2 |
| *Inflows* | 360.4 | 54.6 | 25.8 | 52.3 | 79.0 | 22.5 | 916.9 | 312.4 | 54.3 |

Depending on the magnitude of the net outflows relative to the volume of a specific spatial compartment, *Eq. D1* distinguishes between two cases:

*1)* The net outflows are non-negligible relative to the compartment volume or $\frac{Q_{NET}}{V_A} > 0$. This condition refers to small and moderate sized compartments, such as *C1, C6, C9, K39, K42,* and *ATH* (Table D1). First, because the observed *TP* concentrations suggest that the scenario $C_B > C_A$ rarely holds true (Table D1), the validity of *Eq. D2* could be examined in the following conditions:

*a)* $Q_A > Q_B$*:* This scenario is related to the impact of the spring freshet and/or extreme precipitation events that result in high hydraulic loading into the two embayments. Under these conditions, the mass exchanges are indeed predominantly unidirectional from the nearshore areas to the outer lake.

*b)* $Q_A \approx Q_B$*:* When the water volume exchanges among adjacent segments are approximately equal, then the validity of *Eq. D2* depends on the corresponding *TP* concentrations.

*i)* $C_A \approx C_B$*:* If the two concentrations are also equal, then the two segments are approximately on a steady state. This scenario may represent the conditions frequently experienced in Kempenfelt Bay (*K39→K42*) or the outer area of Cook's Bay and the main basin (*C9→S15*).

*ii)* $C_A > C_B$*:* This scenario reproduces the concentration gradient established at the inner segments of Cook's Bay (*C1→C6→C9*) or the interface between the main basin and Atherley Narrows (*K45→ATH*). Thus, both areas can be assumed to be on par with the patterns postulated by the feedforward model.

*c)* $Q_A < Q_B$*:* This condition mainly refers to wind-induced circulation patterns that predominantly drive water masses to south- or westward directions. In this case, *Eq. D2* is violated and the model used in this analysis cannot describe the day-to-day *TP* variability. In particular, because of the spatial *TP* trends typically experienced in the system, the model fails to account for the dilution effects of the water masses from the outer lake on the two embayments. The latter likelihood also suggests that the net *TP* loss rates presented in the study may overestimate the actual net sedimentation occurring in the nearshore sites.

*2)* The net outflows are negligible relative to the compartment volume or $\frac{Q_{NET}}{V_A} \to 0$. This condition refers to large compartments, such as *S15, K45,* and *E51* (Table D1). In this scenario, the main implication of the feedforward model is that the largest fraction of the inflowing *TP* loads from the watershed and/or antecedent lake segments is being subjected to sedimentation within those compartments, rather than carried forward to the subsequent segments. Notably, this simplification could have led to an overestimation of the sedimentation rates in the eastern basin, if there were substantial exogenous loads and a distinct gradient of *TP* concentrations.

**D2) Analytical solutions of the continuously stirred tank reactor model with sinusoidal loading:**

$W_{avg(i)}$ = mean loading entering segment *i* (tonnes/day)

$W_{amp(i)}$ = amplitude around the mean loading $W_{avg(i)}$ (tonnes/day)

$\lambda_{(i)}$ = total loss rate (day$^{-1}$), where $\lambda_{(i)} = \kappa_{1(i)} + \kappa_{2(i)}$; $\kappa_{1(i)} = \frac{Q_{(i)}}{V_{(i)}}$; $\kappa_{2(i)} = k_{(i)} + \frac{v_{(i)}}{H_{(i)}}$

$Q_{(i)}$ = the volumetric outflow rate for segment *i* (m$^3$/day)

$V_{(i)}$ = volume of segment *i* (m$^3$)

$k_{(i)}$ = segment-specific first-order decay rate (day$^{-1}$)

$v_{(i)}$ = segment-specific settling velocity (m/day)

$H_{(i)}$ = mean depth of segment *i* (m)

$\theta_{(i)}$ = phase shift of the loading from the standard wave (radians),

$\varphi_{(i)}$ = an additional phase shift related to the segment-specific response,

$\omega_{(i)}$ = angular frequency of the loading oscillation (radians/day)

*A-level segments* that initiate a chain of reactors: sites K39, C1, E51

$$C_A = \frac{W_{avgA}}{\lambda_A V_A} + \frac{W_{amp\,A}}{V_A\sqrt{\lambda_A{}^2 + \omega_A{}^2}} \sin[\omega_A t - \theta_A - \phi_A(\omega_A)]$$

where $\omega = \frac{2\pi}{T}$ , $\phi_A(\omega_A) = \arctan(\frac{\omega_A}{\lambda_A})$

*B-level segments* that receive mass from the A-level segments K39 and C1: sites K42, C6

$$C_B = \frac{Q_A W_{avgA}}{\lambda_A V_A \lambda_B V_B} + \frac{W_{avgB}}{\lambda_B V_B} + \frac{W_{amp\,B}}{V_B\sqrt{\lambda_B{}^2 + \omega_B{}^2}} \sin[\omega_B t - \theta_B - \phi_B(\omega_B)]$$

$$+ \frac{Q_A W_{amp\,A}}{V_A V_B \sqrt{\lambda_A{}^2 + \omega_A{}^2}\sqrt{\lambda_B{}^2 + \omega_B{}^2}} \sin[\omega_A t - \theta_A - \phi_A(\omega) - \phi_B(\omega_B)]$$

where $\phi_B(\omega_B) = \arctan(\frac{\omega_B}{\lambda_B})$

*C-level segment* that receives mass from the A-level segment C1 and the B-level segment C6: site C9

$$C_{C9} = \frac{Q_{C6}Q_{C1}W_{avgC1}}{\lambda_{C9}V_{C9}\lambda_{C6}V_{C6}\lambda_{C1}V_{C1}} + \frac{Q_{C6}W_{avgC6}}{\lambda_{C9}V_{C9}\lambda_{C6}V_{C6}} + \frac{W_{avgC9}}{\lambda_{C9}V_{C9}}$$
$$+ \frac{W_{amp\,C9}}{V_{C9}\sqrt{\lambda_{C9}^{\,2} + \omega_{C9}^{\,2}}}\sin[\omega_{C9}t - \theta_{C9} - \phi_{C9}(\omega_{C9})]$$
$$+ \frac{Q_{C6}W_{amp\,C6}}{V_{C9}V_{C6}\sqrt{\lambda_{C6}^{\,2} + \omega_{C6}^{\,2}}\sqrt{\lambda_{C9}^{\,2} + \omega_{C9}^{\,2}}}\sin[\omega_{C6}t - \theta_{C6} - \phi_{C6}(\omega_{C6}) - \phi_{C9}(\omega_{C9})]$$
$$+ \frac{Q_{C6}Q_{C1}W_{amp\,C1}}{V_{C9}V_{C1}V_{C6}\sqrt{\lambda_{C1}^{\,2} + \omega_{C1}^{\,2}}\sqrt{\lambda_{C6}^{\,2} + \omega_{C6}^{\,2}}\sqrt{\lambda_{C9}^{\,2} + \omega_{C9}^{\,2}}}\sin[\omega_{C1}t - \theta_{C1} - \phi_{C1}(\omega_{C1}) - \phi_{C6}(\omega_{C6}) - \phi_{C9}(\omega_{C9})]$$

where $\phi_{C9}(\omega_{C9}) = \arctan(\frac{\omega_{C9}}{\lambda_{C9}})$

*D-level segment* that receives mass from the A-level segment C1, the B-level segment C6, and the C-level segment C9: site S15

$$C_{S15} = \frac{Q_{C9}Q_{C6}Q_{C1}W_{avgC1}}{\lambda_{S15}V_{S15}\lambda_{C9}V_{C9}\lambda_{C6}V_{C6}\lambda_{C1}V_{C1}} + \frac{Q_{C9}Q_{C6}W_{avgC6}}{\lambda_{S15}V_{S15}\lambda_{C9}V_{C9}\lambda_{C6}V_{C6}} + \frac{Q_{C9}W_{avgC9}}{\lambda_{S15}V_{S15}\lambda_{C9}V_{C9}} + \frac{W_{avgS15}}{\lambda_{S15}V_{S15}}$$
$$+ \frac{W_{amp\,S15}}{V_{S15}\sqrt{\lambda_{S15}^{\,2} + \omega_{S15}^{\,2}}}\sin[\omega_{S15}t - \theta_{S15} - \phi_{S15}(\omega_{S15})]$$
$$+ \frac{Q_{C9}W_{amp\,C9}}{V_{S15}V_{C9}\sqrt{\lambda_{C9}^{\,2} + \omega_{C9}^{\,2}}\sqrt{\lambda_{S15}^{\,2} + \omega_{S15}^{\,2}}}\sin[\omega_{C9}t - \theta_{C9} - \phi_{C9}(\omega_{C9}) - \phi_{S15}(\omega_{S15})]$$
$$+ \frac{Q_{C6}Q_{C9}W_{amp\,C6}}{V_{S15}V_{C9}V_{C6}\sqrt{\lambda_{S15}^{\,2} + \omega_{S15}^{\,2}}\sqrt{\lambda_{C9}^{\,2} + \omega_{C9}^{\,2}}\sqrt{\lambda_{C6}^{\,2} + \omega_{C6}^{\,2}}}\sin[\omega_{C6}t - \theta_{C6} - \phi_{C6}(\omega_{C6}) - \phi_{C9}(\omega_{C9}) - \phi_{S15}(\omega_{S15})]$$

$$+\frac{Q_{C1}Q_{C6}Q_{C9}W_{amp\,C1}}{V_{S15}V_{C9}V_{C6}V_{C1}\sqrt{\lambda_{S15}{}^2+\omega_{S15}{}^2}\sqrt{\lambda_{C9}{}^2+\omega_{C9}{}^2}\sqrt{\lambda_{C6}{}^2+\omega_{C6}{}^2}\sqrt{\lambda_{C1}{}^2+\omega_{C1}{}^2}}\sin[\omega_{C1}t-\theta_{C1}-\phi_{C1}(\omega_{C1})-\phi_{C6}(\omega_{C6})-\phi_{C9}(\omega_{C9})-\phi_{S15}(\omega_{S15})]$$

*E-level segment* that receives mass from the A-level segment E51, and the series of reactors from Kempenfelt Bay and Cook's Bay: site K45

$$C_E = C_{K45} + C_{E51} + C_{Kempenfelt\ Bay\_chain} + C_{Cook's\ Bay\_chain}$$

$$C_{K45} = \frac{W_{avgK45}}{\lambda_{K45}V_{K45}} + \frac{W_{amp\,K45}}{V_{K45}\sqrt{\lambda_{K45}{}^2+\omega^2}}\sin[\omega_{K45}t-\theta_{K45}-\phi_{K45}(\omega_{K45})]$$

$$C_{E51} = \frac{Q_{E51}W_{avgE51}}{\lambda_{E51}V_{E51}\lambda_{K45}V_{K45}} + \frac{Q_{E51}W_{amp\,E51}}{V_{E51}V_{K45}\sqrt{\lambda_{E51}{}^2+\omega_{E51}{}^2}\sqrt{\lambda_{K45}{}^2+\omega_{K45}{}^2}}\sin[\omega_{E51}t-\theta_{E51}-\phi_{E51}(\omega_{E51})-\phi_{K45}(\omega_{K45})]$$

$$C_{Kempenfelt\ Bay\_chain} = \frac{Q_{K42}Q_{K39}W_{avgK39}}{\lambda_{K45}V_{K45}\lambda_{K39}V_{K39}\lambda_{K42}V_{K42}} + \frac{Q_{K42}W_{avgK42}}{\lambda_{K45}V_{K45}\lambda_{K42}V_{K42}}$$
$$+\frac{Q_{K42}W_{amp\,K42}}{V_{K45}V_{K42}\sqrt{\lambda_{K42}{}^2+\omega_{K42}{}^2}\sqrt{\lambda_{K45}{}^2+\omega_{K45}{}^2}}\sin[\omega_{K42}t-\theta_{K42}-\phi_{K42}(\omega_{K42})-\phi_{K45}(\omega_{K45})]$$
$$+\frac{Q_{K42}Q_{K39}W_{amp\,K39}}{V_{K45}V_{K39}V_{K42}\sqrt{\lambda_{K39}{}^2+\omega_{K39}{}^2}\sqrt{\lambda_{K42}{}^2+\omega_{K42}{}^2}\sqrt{\lambda_{K45}{}^2+\omega_{K45}{}^2}}\sin[\omega_{K39}t-\theta_{K39}-\phi_{K39}(\omega_{K39})-\phi_{K42}(\omega_{K42})-\phi_{K45}(\omega_{K45})]$$

$$C_{Cook's\ Bay\_chain} = \frac{Q_{S15}Q_{C9}Q_{C6}Q_{C1}W_{avgC1}}{\lambda_{K45}V_{K45}\lambda_{S15}V_{S15}\lambda_{C9}V_{C9}\lambda_{C6}V_{C6}\lambda_{C1}V_{C1}} + \frac{Q_{S15}Q_{C9}Q_{C6}W_{avgC6}}{\lambda_{K45}V_{K45}\lambda_{S15}V_{S15}\lambda_{C9}V_{C9}\lambda_{C6}V_{C6}} + \frac{Q_{S15}Q_{C9}W_{avgC9}}{\lambda_{K45}V_{K45}\lambda_{S15}V_{S15}\lambda_{C9}V_{C9}} + \frac{Q_{S15}W_{avgS15}}{\lambda_{K45}V_{K45}\lambda_{S15}V_{S15}}$$
$$+\frac{Q_{S15}W_{amp\,S15}}{V_{K45}V_{S15}\sqrt{\lambda_{S15}{}^2+\omega_{S15}{}^2}\sqrt{\lambda_{K45}{}^2+\omega_{K45}{}^2}}\sin[\omega_{S15}t-\theta_{S15}-\phi_{S15}(\omega_{S15})-\phi_{K45}(\omega_{K45})]$$
$$+\frac{Q_{C9}Q_{S15}W_{amp\,C9}}{V_{K45}V_{S15}V_{C9}\sqrt{\lambda_{K45}{}^2+\omega_{K45}{}^2}\sqrt{\lambda_{S15}{}^2+\omega_{S15}{}^2}\sqrt{\lambda_{C9}{}^2+\omega_{C9}{}^2}}\sin[\omega_{C9}t-\theta_{C9}-\phi_{C9}(\omega_{C9})-\phi_{S15}(\omega_{S15})-\phi_{K45}(\omega_{K45})]$$
$$+\frac{Q_{C6}Q_{C9}Q_{S15}W_{amp\,C6}}{V_{K45}V_{S15}V_{C9}V_{C6}\sqrt{\lambda_{K45}{}^2+\omega_{K45}{}^2}\sqrt{\lambda_{S15}{}^2+\omega_{S15}{}^2}\sqrt{\lambda_{C9}{}^2+\omega_{C9}{}^2}\sqrt{\lambda_{C6}{}^2+\omega_{C6}{}^2}}\sin[\omega_{C6}t-\theta_{C6}-\phi_{C6}(\omega_{C6})-\phi_{C9}(\omega_{C9})-\phi_{S15}(\omega_{S15})-\phi_{K45}(\omega_{K45})]$$
$$+\frac{Q_{C1}Q_{C6}Q_{C9}Q_{S15}W_{amp\,C1}}{V_{K45}V_{S15}V_{C9}V_{C6}V_{C1}\sqrt{\lambda_{K45}{}^2+\omega_{K45}{}^2}\sqrt{\lambda_{S15}{}^2+\omega_{S15}{}^2}\sqrt{\lambda_{C9}{}^2+\omega_{C9}{}^2}\sqrt{\lambda_{C6}{}^2+\omega_{C6}{}^2}\sqrt{\lambda_{C1}{}^2+\omega_{C1}{}^2}}\times$$

$$\times \sin[\omega_{C1}t - \theta_{C1} - \phi_{C1}(\omega_{C1}) - \phi_{C6}(\omega_{C6}) - \phi_{C9}(\omega_{C9}) - \phi_{S15}(\omega_{S15}) - \phi_{K45}(\omega_{K45})]$$

*F-level segment* that receives mass from the E-level segment K45: site ATH

$$C_F = C_{ATH} + C_{K45} + C_{E51} + C_{Kempenfelt\ Bay\_chain} + C_{Cook's\ Bay\_chain}$$

$$C_{ATH} = \frac{W_{avgATH}}{\lambda_{ATH} V_{ATH}} + \frac{W_{amp\ ATH}}{V_{ATH}\sqrt{\lambda_{ATH}^{\ 2} + \omega_{ATH}^{\ 2}}} \sin[\omega_{ATH}t - \theta_{ATH} - \phi_{ATH}(\omega_{ATH})]$$

$$C_{K45} = \frac{Q_{K45} W_{avgK45}}{\lambda_{ATH} V_{ATH} \lambda_{K45} V_{K45}} + \frac{Q_{K45} W_{amp\ K45}}{V_{K45} V_{ATH}\sqrt{\lambda_{ATH}^{\ 2} + \omega_{ATH}^{\ 2}}\sqrt{\lambda_{K45}^{\ 2} + \omega_{K45}^{\ 2}}} \sin[\omega_{K45}t - \theta_{K45} - \phi_{K45}(\omega_{K45}) - \phi_{ATH}(\omega_{ATH})]$$

$$C_{E51} = \frac{Q_{K45} Q_{E51} W_{avgE51}}{\lambda_{ATH} V_{ATH} \lambda_{K45} V_{K45} \lambda_{E51} V_{E51}} + \frac{Q_{K45} Q_{E51} W_{amp\ E51}}{V_{ATH} V_{E51} V_{K45}\sqrt{\lambda_{E51}^{\ 2} + \omega_{E51}^{\ 2}}\sqrt{\lambda_{K45}^{\ 2} + \omega_{K45}^{\ 2}}\sqrt{\lambda_{ATH}^{\ 2} + \omega_{ATH}^{\ 2}}} \sin[\omega_{E51}t - \theta_{E51} - \phi_{E51}(\omega_{E51}) - \phi_{K45}(\omega_{K45}) - \phi_{ATH}(\omega_{ATH})]$$

$$C_{Kempenfelt\ Bay\_chain} = \frac{Q_{K45} Q_{K42} Q_{K39} W_{avgK39}}{\lambda_{ATH} V_{ATH} \lambda_{K45} V_{K45} \lambda_{K42} V_{K42} \lambda_{K39} V_{K39}} + \frac{Q_{K45} Q_{K42} W_{avgK42}}{\lambda_{ATH} V_{ATH} \lambda_{K42} V_{K42} \lambda_{K45} V_{K45}}$$

$$+ \frac{Q_{K42} Q_{K45} W_{amp\ K42}}{V_{ATH} V_{K45} V_{K42}\sqrt{\lambda_{ATH}^{\ 2} + \omega_{ATH}^{\ 2}}\sqrt{\lambda_{K45}^{\ 2} + \omega_{K45}^{\ 2}}\sqrt{\lambda_{K42}^{\ 2} + \omega_{K42}^{\ 2}}} \sin[\omega_{K42}t - \theta_{K42} - \phi_{K42}(\omega_{K42}) - \phi_{K45}(\omega_{K45}) - \phi_{ATH}(\omega_{ATH})]$$

$$+ \frac{Q_{K39} Q_{K42} Q_{K45} W_{amp\ K39}}{V_{ATH} V_{K45} V_{K42} V_{K39}\sqrt{\lambda_{ATH}^{\ 2} + \omega_{ATH}^{\ 2}}\sqrt{\lambda_{K45}^{\ 2} + \omega_{K45}^{\ 2}}\sqrt{\lambda_{K42}^{\ 2} + \omega_{K42}^{\ 2}}\sqrt{\lambda_{K39}^{\ 2} + \omega_{K39}^{\ 2}}} \sin[\omega_{K39}t - \theta_{K39} - \phi_{K39}(\omega_{K39}) - \phi_{K42}(\omega_{K42}) - \phi_{K45}(\omega_{K45}) - \phi_{ATH}(\omega_{ATH})]$$

$$C_{Cook's\ Bay\_chain} = \frac{Q_{K45} Q_{S15} Q_{C9} Q_{C6} Q_{C1} W_{avgC1}}{\lambda_{ATH} V_{ATH} \lambda_{K45} V_{K45} \lambda_{S15} V_{S15} \lambda_{C9} V_{C9} \lambda_{C6} V_{C6} \lambda_{C1} V_{C1}}$$

$$+ \frac{Q_{K45} Q_{S15} Q_{C9} Q_{C6} W_{avgC6}}{\lambda_{ATH} V_{ATH} \lambda_{K45} V_{K45} \lambda_{S15} V_{S15} \lambda_{C9} V_{C9} \lambda_{C6} V_{C6}} + \frac{Q_{K45} Q_{S15} Q_{C9} W_{avgC9}}{\lambda_{ATH} V_{ATH} \lambda_{K45} V_{K45} \lambda_{S15} V_{S15} \lambda_{C9} V_{C9}} + \frac{Q_{K45} Q_{S15} W_{avgS15}}{\lambda_{ATH} V_{ATH} \lambda_{K45} V_{K45} \lambda_{S15} V_{S15}}$$

$$+ \frac{Q_{K45} Q_{S15} W_{amp\ S15}}{V_{ATH} V_{K45} V_{S15}\sqrt{\lambda_{ATH}^{\ 2} + \omega_{ATH}^{\ 2}}\sqrt{\lambda_{K45}^{\ 2} + \omega_{K45}^{\ 2}}\sqrt{\lambda_{S15}^{\ 2} + \omega_{S15}^{\ 2}}} \sin[\omega_{S15}t - \theta_{S15} - \phi_{S15}(\omega_{S15}) - \phi_{K45}(\omega_{K45}) - \phi_{ATH}(\omega_{ATH})]$$

$$+ \frac{Q_{C9} Q_{S15} Q_{K45} W_{amp\ C9}}{V_{ATH} V_{K45} V_{S15} V_{C9}\sqrt{\lambda_{ATH}^{\ 2} + \omega_{ATH}^{\ 2}}\sqrt{\lambda_{K45}^{\ 2} + \omega_{K45}^{\ 2}}\sqrt{\lambda_{S15}^{\ 2} + \omega_{S15}^{\ 2}}\sqrt{\lambda_{C9}^{\ 2} + \omega_{C9}^{\ 2}}} \sin[\omega_{C9}t - \theta_{C9} - \phi_{C9}(\omega_{C9}) - \phi_{S15}(\omega_{S15}) - \phi_{K45}(\omega_{K45}) - \phi_{ATH}(\omega_{ATH})]$$

$$+\frac{Q_{C6}Q_{C9}Q_{S15}Q_{K45}W_{amp\,C6}}{V_{ATH}V_{K45}V_{S15}V_{C9}V_{C6}\sqrt{\lambda_{ATH}{}^2+\omega_{ATH}{}^2}\sqrt{\lambda_{K45}{}^2+\omega_{K45}{}^2}\sqrt{\lambda_{S15}{}^2+\omega_{S15}{}^2}\sqrt{\lambda_{C9}{}^2+\omega_{C9}{}^2}\sqrt{\lambda_{C6}{}^2+\omega_{C6}{}^2}}\times$$

$$\times\sin[\omega_{C6}t-\theta_{C6}-\phi_{C6}(\omega_{C6})-\phi_{C9}(\omega_{C9})-\phi_{S15}(\omega_{S15})-\phi_{K45}(\omega_{K45})-\phi_{ATH}(\omega_{ATH})]$$

$$+\frac{Q_{C1}Q_{C6}Q_{C9}Q_{S15}Q_{K45}W_{amp\,C1}}{V_{ATH}V_{K45}V_{S15}V_{C9}V_{C6}V_{C1}\sqrt{\lambda_{ATH}{}^2+\omega_{ATH}{}^2}\sqrt{\lambda_{K45}{}^2+\omega_{K45}{}^2}\sqrt{\lambda_{S15}{}^2+\omega_{S15}{}^2}\sqrt{\lambda_{C9}{}^2+\omega_{C9}{}^2}\sqrt{\lambda_{C6}{}^2+\omega_{C6}{}^2}\sqrt{\lambda_{C1}{}^2+\omega_{C1}{}^2}}\times$$

$$\times\sin[\omega_{C1}t-\theta_{C1}-\phi_{C1}(\omega_{C1})-\phi_{C6}(\omega_{C6})-\phi_{C9}(\omega_{C9})-\phi_{S15}(\omega_{S15})-\phi_{K45}(\omega_{K45})-\phi_{ATH}(\omega_{ATH})]$$

## E) STRUCTURAL EQUATION MODEL FOR LAKE SIMCOE

In the SEM practice, the important issue of extracting inference from model results can be classified into three categories: (i) *strictly confirmatory*: a single model is formulated and tested against datasets, ideally after model specification. In this case, the model can be accepted or rejected, (ii) *alternative models*: several prespecified models are tested against single set of data. In this case, one of the models should be selected, and (iii) *model generating*: the analysis starts with a tentative model, which is subjected to evaluation and modification. These respecifications should provide meaningful interpretations and the final model needs further confirmation (Bollen, 1989). Our analysis falls under the third category. In particular, we conducted an exploratory analysis that derived the present structural equation model (Fig. 1c) from an original one that involved more observed variables associated with the planktonic food web dynamics in Lake Simcoe (phosphate, total ammonia, algal biovolume, *Daphnia* sp.). The evaluation of the different model modifications was based on the Deviance Information Criterion (Spiegelhalter et al., 2002) and thus our SEM analysis is founded upon the most parsimonious construct, i.e., the model that more effectively balances between performance and complexity.

Using the classical SEM notation, we present the matrices' forms and the specific assumptions made for the non-recursive structural equation model for Lake Simcoe (Fig. 1b & Fig. 1-SI). The exogenous latent variable measurement model consists of four matrices; i.e. $X$ is a $q$ x 1 vector of observable indicators of the independent latent variables $\xi$; $\Lambda_x$ is a $q$ *x* $n$ matrix of coefficients relating $X$ to $\xi$; $\xi$ is a $n$ x 1 vector of independent (exogenous) latent variables; and, $\delta$ is a $q$ x 1 vector of measurement errors for $X$. In the present model, we included four ($n$ = 4) exogenous latent variables $\xi$ which were described from five ($q$ = 5) indicator variables; i.e., *TN* and *DIN* were used for the latent variable "Nitrogen"; *Secchi Disc Depth* for the latent variable "Light"; *TP* for the latent variable "Phosphorus" and *Temperature* for the respective latent variable. Thus,

$$X = \begin{bmatrix} X_1 = TN \\ X_2 = DIN \\ X_3 = Secchi\ Disc\ Depth \\ X_4 = TP \\ X_5 = Temperature \end{bmatrix}, \Lambda_X = \begin{bmatrix} \lambda_1 & 0 & 0 & 0 \\ \lambda_2 & 0 & 0 & 0 \\ 0 & \lambda_3 & 0 & 0 \\ 0 & 0 & \lambda_4 & 0 \\ 0 & 0 & 0 & \lambda_5 \end{bmatrix}, \xi = \begin{bmatrix} \xi_1 = Nitrogen \\ \xi_2 = Light \\ \xi_3 = Phosphorus \\ \xi_4 = Temperature \end{bmatrix}, \delta = \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \end{bmatrix}$$

The endogenous latent variable measurement model also consists of four matrices; i.e., $Y$ is a $p$ x 1 vector of observable indicators of the dependent latent variables $\eta$; $\Lambda_y$ is a $p$ x $m$ matrix of coefficients relating $Y$ to $\eta$; $\eta$ is a $m$ x 1 vector of dependent (endogenous) latent variables; $\varepsilon$ is a $p$ x 1 vector of measurement errors for $Y$; For our model, two indicator variables ($p$ = 2) were used for the representation of two ($m$ = 2) endogenous latent variables; i.e., *Chlorophyll a* and *Zooplankton* biomass were used as indicators for the latent variables "Phytoplankton" and "Herbivorous Grazing". Thus, the exogenous latent variable measurement model can be described by the four matrices:

$$Y = \begin{bmatrix} Y_1 = Chlorophyll\ a \\ Y_2 = Zooplankton \end{bmatrix}, \Lambda_Y = \begin{bmatrix} \lambda_6 & 0 \\ 0 & \lambda_7 \end{bmatrix}, \eta = \begin{bmatrix} \eta_1 = Phytoplankton \\ \eta_2 = Herbivorous\ Grazing \end{bmatrix}, \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix}$$

The additional three matrices for the latent variable model are:

$$\Gamma = \begin{bmatrix} \gamma_1 & \gamma_2 & \gamma_3 & 0 \\ 0 & 0 & 0 & \gamma_4 \end{bmatrix}, \quad B = \begin{bmatrix} \beta_1 & \beta_2 \end{bmatrix}, \quad \zeta = \begin{bmatrix} \zeta_1 \\ \zeta_2 \end{bmatrix}$$

where $\Gamma$ and $B$ are the matrices of coefficients for the latent exogenous and endogenous variables, respectively; $\zeta$ is the vector of latent (structural) errors. As inferred from the path diagram, the associated covariance matrices of the model, Cov($\xi$) = $\Phi$($n$ x $n$): lake-wide covariances between the independent variables $\xi$; Cov($\varepsilon$) = $\Theta_\varepsilon$($p$ x $p$): segment-specific covariances between the measurement errors in $Y$; Cov$(\delta)$ = $\Theta_\delta$($q$ x $q$): segment-specific covariances between the measurement errors in $X$;

Cov$(\zeta)$ = $\Psi$($m$ x $m$): lake-wide covariances between the structural errors $\zeta$, have the off-diagonal elements equal to zero:

$$\Theta_\varepsilon = \begin{bmatrix} \mathrm{var}(\varepsilon_1) & \\ 0 & \mathrm{var}(\varepsilon_2) \end{bmatrix}, \Theta_\delta = \begin{bmatrix} \mathrm{var}(\delta_1) & & & & \\ 0 & \mathrm{var}(\delta_2) & & & \\ 0 & 0 & \mathrm{var}(\delta_3) & & \\ 0 & 0 & 0 & \mathrm{var}(\delta_4) & \\ 0 & 0 & 0 & 0 & \mathrm{var}(\delta_5) \end{bmatrix},$$

$$\Psi = \begin{bmatrix} \psi_{11} & \\ 0 & \psi_{22} \end{bmatrix},\ \Phi = \begin{bmatrix} \phi_{11} & & & \\ 0 & \phi_{22} & & \\ 0 & 0 & \phi_{33} & \\ 0 & 0 & 0 & \phi_{44} \end{bmatrix}$$

The metric of the latent variables was set by fixing one loading in each column of $\Lambda_X$ and $\Lambda_Y$ to 1.0. In this particular case, we assumed that $\lambda_2 = \lambda_3 = \lambda_4 = \lambda_5 = \lambda_6 = \lambda_7 = 1.0$. The hierarchical Bayesian configuration of the Lake Simcoe SEM for an arbitrary observation $i$ at segment $j$ can be specified as follows:

$$X_{1ij} = \lambda_{1j}\xi_{1ij} + \delta_{1j},\ X_{2ij} = \xi_{1ij} + \delta_{2ij}$$

$$X_{3ij} = \xi_{2ij} + \delta_{3ij}$$

$$X_{4ij} = \xi_{3ij} + \delta_{4ij}$$

$$X_{5ij} = \xi_{4ij} + \delta_{5ij}$$

$$\delta_{.j} \sim N(0, \Theta_{\delta j}),\ \xi_{.} \sim N(0, \Phi)$$

$$Y_{1ij} = \eta_{1ij} + \varepsilon_{1ij}$$

$$Y_{2ij} = \eta_{2ij} + \varepsilon_{2ij}$$

$$\varepsilon_{.j} \sim N(0, \Theta_{\varepsilon j})$$

$$\eta_{1ij} = \gamma_{1j}\xi_{1ij} + \gamma_{2j}\xi_{2ij} + \gamma_{3j}\xi_{3ij} + \beta_{1j}\eta_{2ij} + \zeta_{1ij}$$

$$\eta_{2ij} = \gamma_{4j}\xi_{4ij} + \beta_{2j}\eta_{1ij} + \zeta_{2ij}$$

$$\zeta_{.} \sim N(0, \Psi)$$

We used independent non-informative conjugate gamma priors (0.001, 0.001) for the elements of the matrices $\Theta_{\delta j}^{-1}$, $\Theta_{\varepsilon j}^{-1}$, $\Phi^{-1}$ and $\Psi^{1}$. Effectively "flat" normal prior distributions with means equal to 0 and precisions equal to 0.0001 along with diffuse gamma priors were used for the hierarchical treatment of the structural parameters and the factor loading $\lambda_1$ or $\lambda_{TN}$. MCMC simulation was used as the computation tool implemented in the WinBUGS software (Lunn et al., 2000). We used two chain runs of 50,000 iterations and convergence was assessed using the modified Gelman-Rubin convergence statistic (Brooks and Gelman, 1998). Generally, the sequences converged rapidly (≈ 2,000 iterations), while the summary statistics reported in this study were based on the last 40,000 draws. We also examined the robustness of the parameter statistics in different levels of subsampling of the posterior space (e.g., thin=1, 2, 5, 10, 20, 40). In this case, we found that a thin of 20 is the optimal sampling intensity to maintain the characterization derived from the model updating exercise, while minimizing

the serial correlation of the MCMC samples used. The accuracy of the posterior estimates was inspected by assuring that the Monte Carlo error (an estimate of the difference between the mean of the sampled values and the true posterior mean; see Lunn et al., 2000) for all the parameters was less than 5% of the sample standard deviation. In the context of aquatic ecology, a simple illustration of a Bayesian SEM configuration can be found in Arhonditsis et al. (2006). Finally, we note that the present formulation differs from the spatially-explicit Bayesian SEM developed for the Neuse River Estuary (Arhonditsis et al., 2007a,b) in that (*i*) none of the latent variables is treated as perfectly measured, i.e., all the measurement errors were explicitly considered; (*ii*) the covariance matrices of the exogenous latent variables and the structural errors were common over the entire lake rather than segment-specific.

***References***


Arhonditsis, G.B., C.A. Stow, H.W. Paerl, L.M. Valdes, L.J. Steinberg, and K.H. Reckhow 2007. Delineation of the role of nutrient dynamics and hydrologic forcing on phytoplankton patterns along a freshwater-marine continuum. *Ecological Modelling* 208, 230-246.

Arhonditsis, G.B., H.W. Paerl, L.M. Valdes, C.A. Stow, L.J. Steinberg, and K.H. Reckhow 2007. Application of Bayesian Structural Equation Modelling for examining the Neuse River Estuary (NC, USA) phytoplankton dynamics. *Estuarine Coastal & Shelf Science* 73, 63-80.

Arhonditsis, G.B., C.A. Stow, L.J. Steinberg, M.A. Kenney, R.C. Lathrop, S.J. McBride, and K.H. Reckhow 2006. Exploring ecological patterns with structural equation modelling and Bayesian analysis. *Ecological Modelling* 192, 385-409.

Bollen, K.A. 1989. *Structural Equations with Latent Variables.* Wiley and Sons, New York

Brooks, S.P., and A. Gelman 1998. Alternative methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics* 7, 434-455.

Lunn, D.J., A. Thomas, N. Best, and D. Spiegelhalter 2000. WinBUGS a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics* and *Computing* 10, 325–337.

Spiegelhalter, D., N. Best, B. Carlin, and A. van der Linde 2002. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* 64, 583–639.