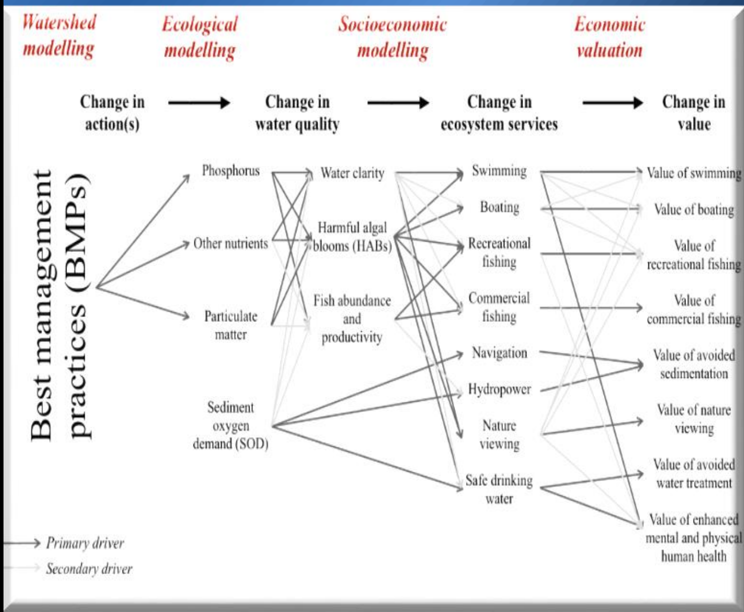
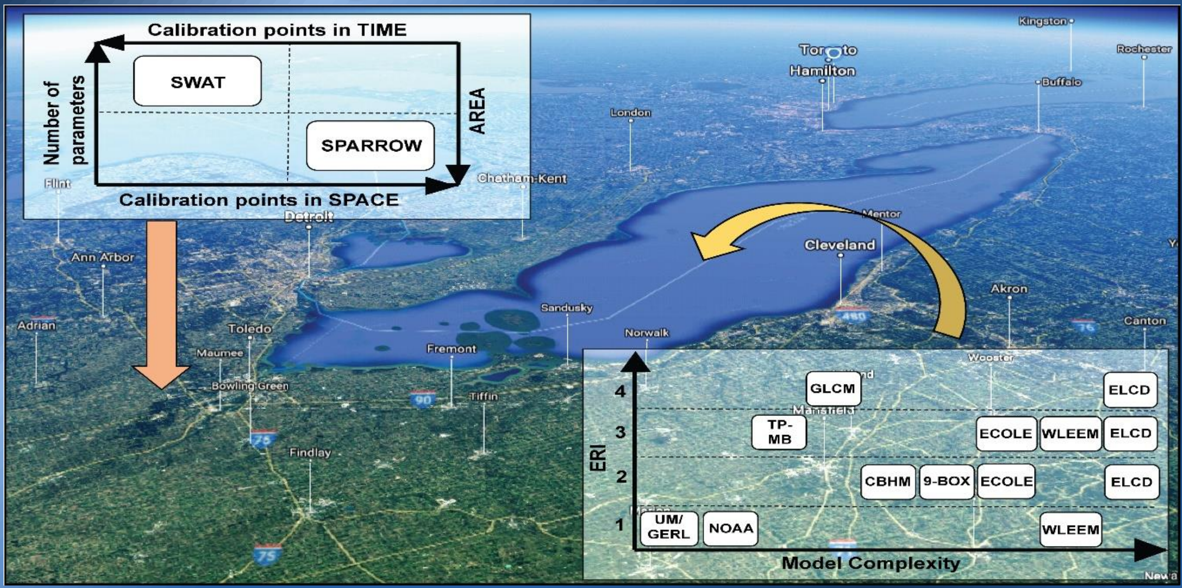


Development of an Integrated Modelling Framework to Guide Adaptive Management Implementation in Lake Erie



Ecological Modelling Laboratory
University of Toronto, Scarborough

Final Report Prepared for:
U.S. Dept. of State, International Joint Commission

Solicitation Number: SAQMMA17R0542

September 2018

Cover images provided by Ecological Modelling Laboratory, University of Toronto Scarborough.

- Upper left: Relationships among human actions, water quality changes, multiple ecosystem goods and services, and associated changes in values in Lake Erie.
- Upper right: An oligotrophic embayment in the Aegean Sea.
- Lower: A suite of watershed and aquatic ecosystem models evaluated in this project.

Work Group Members

Project Lead: Prof. George B. Arhonditsis

Senior Team: Dr Alex Neumann, Dr Yuko Shimoda, Dr Dong-Kyun Kim, Dr Feifei Dong, Dr Gabriela Onandia,

Supporting Team: Dr Vincent Cheng, Ms Cindy Yang, Ms Aisha Javed, Ms Meghan Brady, Ms Ariola Visha, Ms Felicity Ni, Mr Zhuowei Xu, Mr Alexander Lau

Ecological Modelling Laboratory

Department of Physical & Environmental Sciences, University of Toronto,

Toronto, Ontario, Canada, M1C1A4

Table of Contents

List of Figures

List of Tables

Acronym List

1D-CBH - 1-Dimensional Central Basin Hypoxia Model
A2EM - Advanced Aquatic Ecosystem Model
AGNPS - AGricultural Non-Point Source Pollution Model
AnnAGNPS - Annualized AGricultural Non-Point Source pollution model
BFI - Base Flow Index
BMPs - Best Management Practices
CBP - Chesapeake Bay Program
cHABs - Toxic cyanobacteria
Chl *a* - Chlorophyll *a*
CI - Cyanobacteria Index
CN - Curve Number
DLBRM - Distributed Large Basin Runoff Model
DO - Dissolved oxygen
DRP - Dissolved reactive phosphorus
DWSM - Dynamic Watershed Simulation Model
EBC - Eastern Basin Cladophora Model
EcoLE - Ecological model of Lake Erie
EFDC - Environmental Fluid Dynamics Code
ELCOM-CAEDYM - Estuary and Lake Computer Model-Computational Aquatic Ecosystem Dynamics Model
EPIC - Erosion Productivity Impact Calculator
ERI - Ecosystem Response Indicator
EUROSEM - European Soil Erosion Model
GA - Green-Ampt
GLCM - Great Lakes Cladophora Model
GLERL - Great Lakes Environmental Research Laboratory
GLWQA - Great Lakes Water Quality Agreement
GWLF - Generalized Watershed Loading Function
H - Hydrogen
HAB - Harmful Algal Bloom
HBV-INCA - Hydrologiska Byråns Vattenbalansavdelning-Integrated Catchment
HRU - Hydrological Response Unit
HSPF - Hydrological Simulation Program-Fortran
HU - Heidelberg University
IJC - International Joint Commission
K - Soil erodibility
LT - LimnoTech, Inc.
LULC - Land use/Land cover
MEF - Model efficiency
MERIS - Medium Resolution Imaging Spectrometer (satellite instrument)
MIKE SHE - MIKE Système Hydrologique Européen
MODFLOW - MODular 3D Finite-Difference Ground-Water FLOW Model
MODIS - Moderate Resolution Imaging Spectroradiometer (satellite instrument)
MRB3 model - Great Lakes and Upper Mississippi, Ohio, and Red River Basins
MT - Metric tonnes
MUSLE - Modified Universal Soil Loss Equation

N - Nitrogen
NOAA - National Oceanic and Atmospheric Administration (U.S.)
NSE - Nash-Sutcliffe efficiency
O - Oxygen
OH - Ohio
OSU - Ohio State University
P - Phosphorus
PBIAS - Percent biass
PFG - Phytoplankton functional group
PP - Particulate phosphorus
r^2 - Coefficient of determination
RCA - Row-Column Advanced Ecological Systems Modeling Program model
RE - Relative error
RUSLE - Revised Universal Soil Loss Equation
SCS-CN - Conservation Service Curve Number
SeaWiFS - Sea-Viewing Wide Field-of-View Sensor
SEDTRAN - SEDiment TRANsport model
SOD - Sediment oxygen demand
SPARROW - SPAtially Referenced Regressions On Watershed attributes
SRP - Soluable reactive phosphorus
SWAN - Simulating WAves Nearshore
SWAT - Soil and Water Assessment Tool
SWMM - Storm Water Management Model
TAMU - Texas A&M University
TBP - Total bioavailable phosphorus
TEV - Total Economic Value
TN - Total Nitrogen
TP - Total Phosphorus
TPMB - Total Phosphorus Mass Balance Model
UM - University of Michigan
USGS - United States Geological Survey
USLE - Universal Soil Loss Equation
VBS - Vegetative buffer strips
WLEEM - Western Basin Lake Erie Ecosystem Model
ZFG - Zooplankton Functional Group

Key Findings-Highlights

- In Lake Erie, a wide variety of both statistical and mathematical models have been developed to evaluate the relationships among watershed physiography, land use patterns, and phosphorus loading, and to predict the response of the lake to different management actions. Recognizing that these models have different conceptual and methodological strengths, one of the priorities of the local research agenda must be to maintain this diversity and further consolidate the ensemble character of the local forecasting tools.

- The watershed modelling work has been based on five independent applications of the same complex mathematical model, i.e., the Soil and Water Assessment Tool (*SWAT*) model. Each application reflects different assumptions and process characterizations in order to quantify the relative importance of the mechanisms that determine phosphorus loading export from the Maumee River. Thus, this strategy–in principle- captures some of the uncertainties in our understanding of watershed attributes and functioning, and can be classified as a starting model ensemble even though the underlying mathematics used to simulate the study site are the same.

- The five *SWAT* models showed nearly excellent goodness-of-fit against *monthly* flow rates and phosphorus *loading* empirical estimates based on a single downstream station. The *SWAT* modelling work in the Maumee River watershed placed little emphasis on evaluating the robustness of the hydrological or nutrient loading predictions with a finer (*daily*) temporal resolution, and even less so in capturing the impact of episodic/extreme precipitation events.

- In an attempt to delineate high-risk areas (or "hot-spots") with greater propensity for nutrient export and downstream delivery rates, the *SWAT* applications coalesced in their projections and identified higher total phosphorus (*TP*) loading rates from the northwestern and southern parts of the Maumee River watershed, as well as a tendency for higher dissolved reactive phosphorus (*DRP*) export rates at the predominantly agricultural central area. However, we caution that these projections require further ground-truthing by considering multiple sites across the Maumee River watershed to recalibrate the models, and by addressing some of their fundamental discrepancies regarding the fertilizer/manure application rates in the croplands or the spatial drainage of soils.

- Analysis of scenarios with commonly applied (fertilizer reduction, tillage replacement) and less frequent (land-use conversions, wetland/buffer restoration) best management practices (*BMPs*) did not provide strong evidence regarding the likelihood to achieve the March-July phosphorus loading targets of 186 metric tonnes of *DRP* and 860 metric tonnes of *TP*, or 40% reduction from the 2008 loads.

- Overall, our technical analysis suggests that the current modelling work in the Maumee River watershed is not ready yet (*i*) to provide robust predictions regarding the long-term achievability of the phosphorus loading targets, and (*ii*) to evaluate the impact of individual episodic events that can carry significant nutrient loads and presumably modulate the water quality conditions downstream. Critical next steps would be to revisit several influential assumptions (tile drainage, fertilizer/manure application rates, land use/land cover data) and recalibrate the existing *SWAT* applications to capture both baseline and event-flow conditions and daily nutrient concentration (not loading) variability in multiple locations rather than a single downstream site. Another challenging aspect is the proper consideration of legacy *P* (e.g., initialization that accommodates the spatial soil *P* variability, sufficient model spin-up period, parameter specification that reproduces the gradual *P* accumulation in the soils) and its ability to reproduce the critical hydrological and transformation mechanisms modulating the *DRP* loading in the Lake Erie basin. As far as other mechanistic models are concerned, *MIKE SHE* and Storm Water Management Model (*SWMM*) could also be considered due to their ability to simulate channel routing and urban *BPM* scenarios.

- The multi-model ensemble for the Lake Erie itself has been based on a wide range of data-driven and process-based models that span the entire complexity spectrum. Specifically, the models included in the multi-model ensemble strategy were two empirical models, *UM/GLERL* Western Basin *HAB* model and *NOAA* Western Basin *HAB* model, and six process-based models: Total Phosphorus Mass Balance Model (*TPMB*), 1-Dimensional Central Basin Hypoxia Model (*1D-CBH*), Ecological model of Lake Erie (*EcoLE*), Western Basin Lake Erie Ecosystem Model (*WLEEM*), the Estuary and Lake Computer Model-Computational Aquatic Ecosystem Dynamics Model (*ELCOM-CAEDYM*); and the Eastern Basin *Cladophora* Model (*ECB*).

- Consistent with the general trend in the international modelling literature, the performance of the aquatic ecological models in Lake Erie declined from physical, chemical to biological variables. Temperature and dissolved oxygen variability were successfully reproduced, but less so the ambient nutrient levels. Model performance against cyanobacteria was inferior relative to chlorophyll *a* concentrations and zooplankton abundance.

- Our skill assessment analysis is based on a finer resolution in time and space, as opposed to the aggregated spatiotemporal (basin- or lake-wide, seasonal/annual time) scales adopted from the local modelling community, and therefore our goodness-of-fit statistics are less favorable. In principle, the selected coarse scales for evaluating model performance are defensible, as they are consistent with those used for the established nutrient loading targets and water quality indicators in Lake Erie. <u>However, there are compelling technical and management reasons why this practice is problematic and it is recommended to be revisited during the next iteration of the modelling framework.</u>

- General evidence from the international literature suggests that the prediction of the annual *TP* load reduction needed (1,130-3,010 MT) to decrease in half the *summer average chlorophyll a concentration* in the western basin is undoubtedly in the right direction. Likewise, the coupling of empirical and process-based models offers a healthy foundation to evaluate competing hypotheses and support forecasts regarding the achievability of the target related to the *maximum 30-day average cyanobacteria biomass* in the same basin. To further reduce the predictive uncertainty, it is important to improve our understanding (and subsequently the representation with the existing process-based models) of several facets of phytoplankton ecology, such as the postulated degree of reliance of phytoplankton growth upon internal nutrient sources (e.g., microbially mediated regeneration, dreissenid or zooplankton excreted material in nearshore and offshore waters, respectively), the internal *P* loading from the sediments, the role of nitrogen, and the trophic interactions with zooplankton.

- We are particularly skeptical about the optimistic predictions of the extent and duration of hypoxia, given our limited knowledge of the sediment diagenesis processes in the central basin and the lack of data related to the vertical profiles of organic matter and phosphorus fractionation or sedimentation/burial rates. There is a rich research agenda that should be put in place with the next iteration of the adaptive management cycle, before we are in a position to predict the degree and timing of the sediment response or the likelihood of unexpected feedback loops that could delay the realization of the anticipated outcomes. The same is true about our ability to forecast the control of *Cladophora* growth in the eastern basin by *P* load reductions. The management efforts will greatly benefit from a high-resolution monitoring of the nearshore zone to provide critical information regarding the causal linkages between the abiotic conditions (e.g., *SRP*, light, temperature) in the surrounding environment and the internal *P* content and sloughing rates in *Cladophora* mats.

- Viewing ecosystems as providers of economically valuable benefits to humans, the concept of ecosystem services effectively links their structural and functional integrity with human welfare. Given the presence of a wide array of feedback loops, ecological unknowns, and other external stressors (i.e., internal loading, dreissenid mussels, different trends between *TP* and *DRP* loading, changing climate and increased frequency of extreme events), we strongly recommend the development of a rigorous framework that quantifies the socioeconomic benefits from a well-functioning ecosystem. <u>This may be proven to be a critically important strategy to gain leeway and keep the investments to the environment going, especially if the water quality improvements in Lake Erie are not realized in a timely manner.</u>

- Consistent with our criticism regarding the skill assessment of the existing modelling work against aggregated spatiotemporal resolution, we also question the adequacy of the coarse scales selected to establish nutrient loading targets and ecosystem response indicators in Lake Erie. This strategy is neither reflective of the range of spatiotemporal dynamics typically experienced in the system nor does it allow to evaluate our progress with ecosystem services at the degree of granularity required to assess the public sentiment. <u>It would seem paradoxical to expect a single-valued standard, based on monitoring and modelling of offshore waters, to capture the water quality conditions in nearshore areas of high public exposure (e.g., beaches). The degree of public satisfaction is primarily determined by the prevailing conditions at a particular recreational site and given date, and less so by the average water quality over the entire basin (or lake) and growing season.</u>

- In the context of adaptive management implementation, we believe that two critical next steps involve the determination of appropriate metrics and scales of expression along with the design of a monitoring program that will allow to effectively track the progress of the system in both time and space.

- From a management standpoint, our technical analysis concludes that the existing modelling work in Lake Erie has advanced significantly our understanding of the major causal linkages/ecosystem processes underlying the local water quality problems. The results of the forecasting exercises presented in the recent literature are pointing to the correct direction with respect to the on-going management efforts. However, we emphatically caution that there are several highly uncertain predictions and management recommendations (e.g., achievability of loading targets, alleviation of hypoxia, and likelihood to control of *Cladophora* growth) that should be critically reviewed through the iterative monitoring-modelling-assessment cycles of adaptive management.

Executive Summary

Adaptive implementation (or "learning while doing") is often considered the only defensible environmental management strategy, as it promotes an iterative implementation process to deal with the uncertainty of ecological forecasts and to mitigate the impact of inefficient management plans. Environmental models are one of the pillars of the adaptive management process, whereby the initial forecasts of management actions are augmented with post-implementation monitoring and the resulting integration of monitoring and modelling provides the basis for revised management actions. In Lake Erie, a unique combination of data-driven and process-based models have been developed to evaluate the relationships among watershed physiography, land use patterns, and phosphorus loading, to elucidate ecological interactions, to understand the mechanisms underlying specific facets of the ecosystem functioning (cyanobacteria dominance, re-engineering of the nearshore zone induced by dreissenid mussels, *Cladophora* proliferation), and to predict the response of the lake to external nutrient loading reductions. Consistent with the scientific process of progressive learning, the present study offers a technical analysis to assist with the design of the next iteration of the modelling framework by identifying strengths and weaknesses of the existing work and pinpointing essential structural augmentations and research/monitoring priorities in order to integrate watershed and aquatic ecosystem processes.

Model ensembles and adaptive management implementation

Recognizing that there is no true model of an ecological system, but rather several adequate descriptions of different conceptual basis and structure, the adoption of a multi-model framework is specifically designed to address the uncertainty inherent in the model selection process. Nonetheless, the presence of multiple models on its own cannot ensure that the decision-making process is reliably supported, as there are several methodological steps required in order (*i*) to identify the conceptual or structural differences of the existing models, and thus determine the actual diversity (or rule out the likelihood of *"pseudo-replication"*) collectively characterizing the multi-model ensemble; (*ii*) to determine the most suitable calibration/validation domain and resolution for evaluating model performance in time and space; and (*iii*) to establish an optimal weighting scheme in order to assign weights to individual models, when integrating over their corresponding predictions, and subsequently determine the most likely outcome along with the associated uncertainty bounds.

In terms of model diversity, the local watershed modelling work has been based on five applications of the same complex mathematical model, i.e., the Soil and Water Assessment Tool (*SWAT*) model, from groups affiliated with Heidelberg University (*HU*), LimnoTech (*LT*), Ohio State University (*OSU*), Texas A&M University (*TAMU*), and the University of Michigan (*UM*). Each application represents an independent attempt to quantify the relative importance of the processes that determine phosphorus loading export from the Maumee River. In doing so, the five *SWAT* models collectively captured some of the uncertainty in our understanding of watershed attributes and functioning, and thus this strategy (hereafter referred to as *SWAT*-ensemble) can be classified as a model ensemble even though the underlying mathematics used to simulate the study site were the same. The *SWAT*-ensemble offered a platform to examine different land-use management scenarios after consultation with agricultural and conservation stakeholders. Except from the *SWAT* applications, an empirical (SPAtially Referenced Regression On Watershed attributes or *SPARROW*) model was also used in a post-hoc delineation exercise to validate the predicted locations with higher propensity of phosphorus export from the Maumee River watershed, but not as a forecasting device to examine the efficiency of different land use management scenarios.

The multi-model ensemble for the Lake Erie itself has been based on a wide range of data-driven and process-based models to examine the achievability of different environmental targets under various external nutrient loading conditions. Specifically, the models included in the multi-model ensemble strategy were two empirical models, *UM/GLERL* Western Basin *HAB* model and *NOAA* Western Basin *HAB* model, and six process-based models: Total Phosphorus Mass Balance Model (*TPMB*), 1-Dimensional Central Basin Hypoxia Model (*1D-CBH*), Ecological model of Lake Erie (*EcoLE*), Western Basin Lake Erie Ecosystem Model (*WLEEM*), the Estuary and Lake Computer Model-Computational Aquatic Ecosystem Dynamics Model (*ELCOM-CAEDYM*); and the Eastern Basin *Cladophora* Model (*ECB*). Regarding the diversity of the process-based modelling work for Lake Erie, the models developed come from the entire complexity spectrum.

SWAT-based ensemble strategy for the Maumee River watershed

Based on pre-specified performance standards for the simulated flow and phosphorus loading rates, the five *SWAT* models showed nearly excellent goodness-of-fit against measured *monthly* flow rates and phosphorus loading empirical estimates. However, none of the existing models reported goodness-of-fit against measured phosphorus concentrations. There is evidence either from the reported fit statistics on a monthly scale (i.e., underestimated flow rates combined with overestimated phosphorus loading and vice-versa) or a few graphs presented in the literature that the existing *SWAT* models (e.g., *HU* and *TAMU*) are characterized by significant bias of the corresponding concentrations, which in turn casts doubt on the process characterization of the simulated phosphorus cycles.

The *SWAT* modelling work in the Maumee River watershed placed little emphasis on evaluating the robustness of the hydrological or nutrient loading predictions with a finer temporal resolution. In an independent error assessment of the *daily* outputs from the *UM* model, we showed that the *UM SWAT* consistently underestimated the flow rates in 20 out of 22 episodic precipitation events (> 1000 $m^3 s^{-1}$), reinforcing the point that the previously reported excellent goodness-of-fit with a coarser (seasonal or monthly) resolution may simply stem from multiple daily errors/biases that cancel each other out when seasonally or monthly averaged, and thus does not necessarily guarantee acceptable performance against finer time scales.

Following the development of a spatially distributed model, the identification of high-risk areas (or "hot-spots") with greater propensity for nutrient export and downstream delivery rates is an important exercise. In the Maumee River watershed, given that the calibration of all five *SWAT* applications was based on a single downstream station, two factors could be responsible for the (dis)agreement among the corresponding delineations: (*i*) the discrepancies among the assumptions made or input data used during the spatial configuration (e.g., tile drainage, fertilizer/manure application rates, land use/land cover or *LULC* data) of the individual models; and (*ii*) the differences in the characterization of processes pertaining to the simulated water and nutrient cycles. Bearing these two major sources of uncertainty in mind, there was a general agreement among the models in identifying higher *TP* loading rates from the northwestern and southern parts of the Maumee River watershed, whereas a tendency for higher dissolved reactive phosphorus (*DRP*) export rates was mainly projected on the predominantly agricultural central area.

In the Maumee River watershed, several scenarios of best management practices (*BMPs*) were designed after considering issues related to their practical implementation and policy feasibility, i.e., commonly applied (fertilizer reduction, tillage replacement) versus less frequent management practices (land-use conversions, wetland/buffer restoration); the ability of *SWAT* to examine certain agricultural activities; and extensive consultation with agricultural and conservation stakeholders. Overall, there was no strong evidence regarding the likelihood to achieve the March-July phosphorus loading targets of 186 metric tonnes of *DRP* and 860 metric tonnes of *TP*, or 40% reduction from the 2008 loads, across the different *BMP* scenarios examined. The attainability of the *TP* loading threshold seems to be more possible relative to the one for *DRP* loading. It is also worth noting that the forecasts associated with commonly applied *BMP* scenarios were somewhat more conservative, in comparison with scenarios that are less frequently applied. The degree of divergence of the individual model forecasts for a given *BMP* scenario examined (or the forecasting spread) offered insights that can meaningfully inform the environmental policy analysis process. Specifically, the forecasting spread increases significantly with the degree of deviation of *BMP* scenarios from the present conditions. The existing *SWAT* applications suggest that for every 50 metric tonnes of reduction achieved the standardized forecasting spread, or the deviation of the five models divided by their corresponding averaged prediction for a given scenario, increases by 1.5% and 13% for *TP* and *DRP*, respectively. To put it another way, the mean forecasted range with the existing loading targets of 860 tonnes for *TP* and 186 tonnes for *DRP* is 635-1085 tonnes and 120-245 tonnes, respectively.

Multi-model ensemble strategy for Lake Erie

In reviewing the pertinent literature, a first notable finding is that the modelling work for Lake Erie has closely followed the recommended methodological protocol when developing models intended to assist environmental management. Following the evolution of each model over time, we can find detailed sensitivity analysis exercises and goodness-of-fit statistics against a wide range of multi-year conditions and spatial domains. On the other hand, because of the complexity of the existing mechanistic models in Lake Erie, the rigorous quantification of their uncertainty can be particularly challenging (if at all possible) and thus has not received substantial attention. Consistent with the general trend in the international modelling literature, the performance of the aquatic ecological models in Lake Erie declined from physical, chemical to biological variables. Specifically, the temperature and *DO* variability were successfully reproduced, and less so the ambient *TP*, *DP*, NO_2+NO_3, and NH_4 levels. Model performance against cyanobacteria was distinctly worse relative to chlorophyll *a* concentrations and zooplankton abundance.

Regarding our skill assessment analysis, it is important to note that our study is based on a point comparison in time and space, as opposed to the aggregated spatiotemporal (basin- or lake-wide, seasonal/annual time) scales adopted from the local modelling community, and therefore our goodness-of-fit statistics are less favorable compared to those reported against coarser spatio-temporal scales. In principle, the selected coarse scales for evaluating model performance in time and space are defensible, as they are consistent with those used for the established nutrient loading targets and water quality indicators in Lake Erie. Nonetheless, given that the majority of these models are based on daily (or sub-daily) simulations within one- to three-dimensional spatial domains, it would seem that the bar of what constitutes an *acceptable* model performance has been lowered significantly. There are compelling reasons why this practice is problematic and should be revisited during the next iteration of the modelling framework. From a technical standpoint, evaluating model goodness-of-fit with a coarser resolution not only entails the risk to obfuscate multiple daily or location-specific errors/biases that cancel each other out when seasonally or spatially averaged, but may also detract the attention from the much-needed critical evaluation of the process characterizations derived after the calibration of (prone-to-overfitting) complex models. In particular, many of the assumptions made or parameter values assigned could be adequate to describe spatially or temporally aggregated patterns, but could also be the culprits for the misrepresentation of important aspects of the intra- or inter-annual and spatial variability (e.g., magnitude of the spring freshet, timing of algal blooms, and response of the nearshore zone to extreme precipitation events). Our independent model-fit reassessment exercise on a daily scale reinforced the importance of the latter issue by showing the distinctly inferior performance of both watershed and aquatic ecosystem models, as well as their inability to capture critical short-term or event-based facets of the simulated terrestrial and aquatic biogeochemical cycles.

Load-Response Curves and Sources of Uncertainty: After forcing the different lake models with a series of nutrient loading reduction scenarios, the likelihood to achieve four ecosystem response indicators (*ERI*s) was examined by the construction of load-response curves. Our assessment of the lessons learned and important issues identified from this forecasting exercise are as follows:

- *Basin-specific overall phytoplankton biomass, represented by summer average chlorophyll a concentration*: The diversity of the existing modelling work in Lake Erie as well as the general evidence from the international literature in terms of our ability to predict total phytoplankton biomass suggest that this forecasting exercise has a lot of potential to meaningfully assist the local management efforts. The multi-model ensemble predicted that the annual *TP* loads into the western basin needed to bring about a 50% decrease in the maximum Chl *a* concentrations range between 1,130 and 3,010 MT.

- *Western basin cyanobacteria biomass, represented by the maximum 30-day average cyanobacteria biomass*: Considering the challenges with the modelling of individual phytoplankton functional groups, the forecasting exercise regarding the likelihood of cyanobacteria blooms under different loading regimes is as robust as it can be realistically expected. The coupling of empirical and process-based models for this *ERI* offers a healthy foundation to evaluate competing hypotheses and advance our knowledge on the suite of factors that may trigger cyanobacteria dominance in Lake Erie. We caution though that the reported range of cumulative Maumee March–July annual loads of 1679–2170 MT for achieving the cyanobacteria harmful algal bloom (c*HAB*) target is likely narrow and does not reflect the actual uncertainty with this *ERI*.

- *Central basin hypoxia, represented by number of hypoxic days; average extent of hypoxic area during summer; and average hypolimnion DO concentration during August and September:* The load-response curves were suggestive of a fairly wide uncertainty range, 2,600–5,100 MT, within which the targeted hypolimnetic *DO* threshold of 4.0 mg L^{-1} can be realized. Likewise, a load reduction anywhere between 3,415-5,955 MT was projected to reduce the average hypoxic extent to 2000 km^2 and the number of hypoxic days between 9 to 42 days. Generally, our study casts doubt on the ability of the existing models to support reliable predictions regarding the likelihood to alleviate the hypoxia in the central basin of Lake Erie, given that our mechanistic understanding of sediment diagenesis, i.e., the characterization of organic matter mineralization and redox-controlled processes within different sediment layers, is still inadequate. There is a rich research agenda that should be in place with the next iteration of the adaptive management cycle, before we are in a position to predict the degree and timing of the sediment response or the likelihood of unexpected feedback loops that could delay the realization of the anticipated outcomes.

- *Eastern basin Cladophora represented by dry weight biomass and stored P content:* The achievability of a threshold value of 30 g dry weight biomass/m^2 was assessed with load–response curves generated in an area centered on the Grand River and covering 40 km of the northern shoreline area out to the 15 m depth contour. The predictions drawn suggested that *P* load reductions will bring about minor decline in the *Cladophora* biomass in the eastern basin. There are three compelling reasons why the modelling of *Cladophora* in the eastern basin still carries little predictive value. First, the domain within which *Cladophora* growth could be regulated by soluble reactive phosphorus (*SRP*) concentrations is extremely low (0.2–1.0 μg P L^{-1}), while year-to-year variability even on the order of 1 μg P L^{-1} could result in variations of depth-integrated biomass by a factor of 3.5. Second, except from the supply by dreissenid excreta, the *SRP* nearshore concentrations are also modulated by the inflows from the Grand River as well as the nearshore-offshore exchanges. Frequent upwelling events driven by favorable winds of 5-10 days period can easily increase *P* supply above saturation levels. Third, although plausible explanations on the factors that accelerate the sloughing rates and their development within the *Cladophora* mats do exist, the mathematical representation of the associated processes is far from adequate. Rather than increasing the complexity of (already) over-parameterized models, the management efforts will be better supported by the development of two empirical models offering causal linkages between the abiotic conditions (e.g., *SRP*, light, temperature) in the surrounding environment and the internal *P* content and sloughing rates in *Cladophora* mats.

Next steps towards the development of an integrated modelling framework

The recommended next steps and outstanding questions/challenges that need to be considered with the design of the next iteration of the modelling framework in Lake Erie are as follows:

- *Persistence of DRP loading from the Maumee River watershed:* Empirical and modelling evidence suggests an increasing *DRP* loading trend after the mid-1990s, which has been attributed to the increased frequency of storm events, excessive fertilizer application rates and timing, and management practices that appear to increase phosphorus accumulation at the soil surface. Given this emerging evidence, one challenging aspect for the evaluation of scenarios with the *SWAT*-ensemble is the proper consideration of legacy *P* (e.g., initialization that accommodates the spatial soil *P* variability, sufficient model spin-up period, parameter specification that reproduces the gradual *P* accumulation in the soils) and its ability to reproduce the critical hydrological and transformation mechanisms modulating the *DRP* loading in the Lake Erie basin.

- *Characterization of watershed processes associated with the nitrogen cycle:* Contrary to our understanding of *TP* fate and transport, a greater proportion of total nitrogen (*TN*) is found in the dissolved phase due to relatively high solubility of nitrogen species, such as nitrite and nitrate, and can be transported by both overland and subsurface flow paths (greater than phosphate due to immobilization of phosphate by clay and other soil chemical constituents). It is thus important to improve our understanding of the watershed processes associated with the nitrogen cycle, given that it could be one of the regulatory factors of the downstream water quality conditions; especially the algal community composition.

- *Impact of extreme precipitation events:* The characterization of surface runoff and subsurface processes during flow events is largely unknown in the area, and therefore the design of high frequency, event-based, water quality sampling coupled with water stable isotope analysis (^{18}O and ^{2}H) should be one of the priorities in our efforts to rectify the misrepresentation of extreme flow conditions.

- *BMP uncertainty:* Although *BMP* implementation is typically based on the stipulation that both short- and long-term effectiveness are guaranteed, emerging evidence is suggestive of moderate water quality improvements in many watersheds and broad variability in their performances, often much lower compared to the specs of the original design from *BMP* experimental studies. As a first step to accommodate *BMP* uncertainty, we thus propose a moderate enhancement with a stochastic time-invariant representation of *BMP* effectiveness in watershed models, followed by the introduction of time-variant probability distributions for *BMP* life-cycle performance. The proposed stochastic augmentation would allow studying the uncertainty of *BMP* scenarios with Monte Carlo simulations, thereby providing a pragmatic tool to assess the likelihood of the achievability of the proposed nutrient loading reduction goals.

- *Need for other process-based models in the Maumee River watershed: SWAT* could be complemented by the modules of other watershed models, especially for surface runoff, groundwater and sediment erosion processes. For the hydrological and sediment processes, *MIKE SHE* seems to be more up-to-date with respect to the mechanisms considered, assuming that local empirical knowledge is available to constrain the additional parameters. Regarding the simulated *P* cycle, *SWAT* has the advantage to explicitly simulate the daily plant growth, but it could be further improved by adopting a dynamic *P* equilibrium concentration. *MIKE SHE* and Storm Water Management Model (*SWMM*) are superior to *SWAT* in channel routing because of their capability to simulate pipe flows. *SWAT* is more suitable for agricultural *BMP*s (e.g., terracing, contouring, strip cropping, tillage operations, crop rotations, and fertilizer application), while the urban *BMP* modules in *SWMM* (e.g., rain gardens, green roofs, infiltration trenches, permeable pavement, and vegetative swales) offer a more reliable alternative model. More importantly, we believe that greater insights will be gained by revisiting several influential assumptions (tile drainage, fertilizer/manure application rates, *LULC* data) and recalibrating the existing *SWAT* applications to capture both baseline and event-flow conditions and daily nutrient concentration (not loading) variability in multiple locations rather than a single downstream site.

- *Empirical modelling in the Lake Erie basin:* Together with the process-based modelling work in the Maumee River watershed, it is also critical to have simpler empirical models in place that not only provide predictive statements confined within the bounds of data-based parameter estimation, but also to constrain processes/fluxes parameterized by mechanistic models or even to validate the corresponding forecasts drawn. An appealing alternative could be the development of a Great Lakes *SPARROW* that narrows the focus of the original model, while maintaining its "global" character. Importantly, the rigid common parameter estimates over the entire spatial model domain can be relaxed by the use of a hierarchical structure that allows to estimate watershed-specific parameters, and thus accommodate the spatial variability within the Great Lakes basin. In addition, rather than the strict data censoring currently implemented, the *SPARROW* practice should become more inclusive. In particular, the calibration datasets could be coupled with measurement-error submodels to characterize our degree of confidence on their quality or to accommodate the serial correlation among nested subwatersheds. This is an important project that will consolidate the presence of an empirical modelling tool to guide the delineation of nutrient hot-spots alongside the process-based modelling work.

- *Improving the credibility of the load-response curves:* Our study highlighted several important structural augmentations of the existing modelling tools that could increase both their heuristic and predictive values as long as commensurate empirical knowledge to constrain the mathematics becomes available from Lake Erie. If we strive to establish predictive linkages between the magnitude and timing of the response of the sediments and different loading regimes, the study of the sediment diagenesis processes is essential in understanding the control of redox chemistry on the vertical profiles of biodegradable organic matter and *P* binding forms. Empirical information is also needed to constrain the submodels/differential equations related to dreissenids, *Cladophora*, and zooplankton. While decent progress has been made in representing the role of dreissenid mussels in the system, little work has been done to properly adapt the existing *Cladophora* submodel to the nearshore zone and even less so to depict the phytoplankton-zooplankton interactions in Lake Erie. Likewise, with the shift in focus to the average conditions of the offshore waters, the nearshore zone has received less attention from the existing modelling work in Lake Erie.

- *Understanding the factors triggering HAB formation:* One important lesson learned from both mechanistic and data-driven models was that both the dissolved reactive and particulate fractions of *TP* load must be taken into account when setting *HAB*-related load targets. Existing empirical estimates show significant variability of the bioavailable fraction of particulate phosphorus (20-45%) in the Maumee River, and several mechanisms (e.g., microbial mineralization, anoxic release from the sediments) could potentially determine the bioavailability of the inflowing material from the time of entry in early spring until the mid-summer initiation of *Microcystis* blooms. Another interesting finding from the modelling work in Lake Erie is that its susceptibility to *HAB* occurrence could be increasing, and this trend could be attributed to changing meteorological conditions, such as warmer temperatures and calmer summer conditions, presence of an increasing reservoir of *Microcystis* seed colonies, and the selective filtering of dreissenids on competing phytoplankton species. Another factor that has received little attention is the importance of the inter-specific competition for various nitrogen forms; in particular, urea and ammonium are considered energetically favorable forms for protein synthesis and therefore predominant stimulants of *Microcystis* blooms. There is emerging evidence from other locations around the Great Lakes of a strong relationship between nitrogen concentration and toxin-producing *Microcystis* strains or microcystin production, but limited empirical or modelling work has been done to evaluate this hypothesis in Lake Erie.

- *Valuation of ecosystem services in Lake Erie:* Given that environmental policy affects both the ecosystem state and the provision of services that human societies benefit from, we argue that the efficacy of the local restoration efforts will be significantly enhanced by the development of a rigorous framework that quantifies the economic benefits from a well-functioning ecosystem. Economic values of ecosystem services can help policy-makers determine the optimal degree of investment and action needed at each time step by defining the monetary trade-offs from different courses of management action. At the beginning of each restoration effort, the total returns and benefits are typically commensurate with the costs and investments, but this pattern may not hold true after a certain point, where we get diminishing (and ultimately negative) returns and marginal benefits. Viewed it from this perspective, it is important to delve into (somewhat underappreciated) ideas, such as the total economic value (*TEV*) of an ecosystem, the degree of our knowledge of the monetary value of ecosystem services in Lake Erie, and the mismatch between the scales where environmental goals are being set and the spatiotemporal domain that predominantly influences the perception of the public.

- *Environmental criteria/standards*: Consistent with our criticism regarding the skill assessment of the existing modelling work against aggregated spatiotemporal (seasonal/annual, basin- or lake-wide) resolution, we also question the adequacy of the coarse scales selected to establish nutrient loading targets and water quality indicators in Lake Erie. This strategy is neither reflective of the range of spatiotemporal dynamics typically experienced in the system nor does it allow to evaluate our progress with ecosystem services at the degree of granularity required to assess the public sentiment. It would seem paradoxical to expect a single-valued standard, based on monitoring and modelling of offshore waters, to capture the water quality conditions in nearshore areas of high public exposure (e.g., beaches). The degree of public satisfaction is primarily determined by the prevailing conditions at a particular recreational site and given date, and less so by the average water quality over the entire basin (or lake) and growing season. In our view, the problems with the outdated practice to basing the water quality assessment on the offshore zone with a coarse time scale are twofold: (*i*) we cannot effectively track the progress with the response of the system, as it is not clear to what extent an incremental improvement in the open waters is translated into distinct changes in the nearshore; and (*ii*) the environment targets and decisions are implicitly disconnected with our aspiration to protect ecosystem services and gauge public satisfaction at the appropriate resolution.

- *Future monitoring:* In the context of adaptive management implementation, we believe that the critical next steps involve the determination of appropriate metrics and scales of expression along with the design of a monitoring program that will allow to effectively track the progress of the system in both time and space. Depending on the *ERI* considered, there are different areas for future augmentation in order to more comprehensively monitor the response of Lake Erie. In particular, the assessment of the trophic status may be more appropriate to revolve around extreme (or maximum allowable) phytoplankton or *TP* levels and must explicitly accommodate all the sources of uncertainty by permitting a realistic frequency of violations. Rather than any type of data averaging, we advocate the assessment of compliance against the proposed probabilistic criteria using daily snapshots collected regularly from different sites during the growing season. The development of the "cyanobacteria index" is certainly useful, but given the technical limitations of the satellite images, we also need other *cHAB* proxy variables that will be collected regularly from the system, including toxins (e.g., Microcystin-LR). The established thresholds for drinking water (1.5 $\mu g\ L^{-1}$) and recreational purposes (20 $\mu g\ L^{-1}$) offer easily defensible targets to track the frequency of compliance of Lake Erie in time and space. Regarding the hypoxia and *Cladophora ERI*s, given our limited mechanistic and quantitative understanding of the primary driving factors, we also propose the development of systematic records for variables that represent direct causal factors of the actual problem, such as phosphorus content in the *Cladophora* tissues, characterization of the organic matter and phosphorus fractionation in the sediments, are the most prudent strategy to move forward.

- *Role of complex mathematical models in Lake Erie*: From a management standpoint, it is important to note that the complex mechanistic models are an absolutely worthwhile activity and will continue to assist the on-going management efforts in a meaningfully way. Even if the structure of complex mathematical models reduces their predictive power or even the ability to conduct rigorous uncertainty analysis, they still offer excellent platforms to gain insights into the direct, indirect, and synergistic effects of the ecological mechanisms forming the foundation of system behavior. For example, the virtual *3D* environment created by *ELCOM-CAEDYM* and/or *WLEEM* can offer a convenient platform to reconcile the coarse-scale (practically offshore) predictions, required to assess the *ERI* achievability, with the granularity that necessitates to elucidate nearshore processes and associated ecosystem services. Being an integral part of the iterative monitoring-modelling-assessment cycles, the foundation of the mechanistic modelling work in Lake Erie can be optimized through reduction of the uncertainty of critical ecological processes or refinement of their structure (e.g., mathematical reformulation of highly sensitive terms, exclusion of irrelevant mechanisms and inclusion of missing ones), thereby enhancing their ability to support ecological forecasts. It is thus critical that one of the priorities of the research agenda should be to maintain the ensemble character of the modelling work in Lake Erie. The wide variety of models that have been developed to understand the major causal linkages/ecosystem processes underlying the local water quality problems are a unique feature that should be embraced and further augmented.

"….Models serve as expressions of ecological understanding, as engines for deductive inference, and as articulations of resource response to management and environmental change. They help bring together scientists, managers, and other stakeholders in a joint assessment of what is known about the system being managed, and facilitate an interdisciplinary approach to understanding through monitoring and assessment…."

William et al. (2009) Adaptive Management: The U.S. Department of the Interior Technical Guide.

1 Introduction

In the context of natural resource management, the central goals of policy analysis aim to identify the important drivers of environmental degradation, to elucidate the sources of controversy, and to put the necessary tools in place in order to anticipate the unexpected. Environmental problems have a way of resurfacing themselves and are rarely (if ever) solved completely. However, even if certain facets of a management problem change over time, the core issues remain the same, and thus it is critical to establish a framework that ensures both continuity in the decision-making process but also iterative adjustments to accommodate the extrinsic non-stationarity or intrinsic stochasticity (Allen et al., 2011; Williams et al., 2011). Viewed from this context, adaptive implementation is a pragmatic strategy that not only acts as a hedge against the ubiquitous uncertainty surrounding the study of environmental systems, but also paves the way for the dual pursuit of management and learning (Walters and Holling, 1990). Adaptive management offers flexibility in making decisions in the face of uncertainty, as scientific learning progressively advances from research, monitoring, and impartial evaluation of the outcomes of past and on-going management actions (Fig. 1). Even though its core principles are often misconstrued as a "trial-and-error" process, adaptive management promotes a *learning-while-doing* mindset whereby policies or operations can be updated accordingly (Williams et al., 2009; Lyons et al., 2010; Conroy et al., 2011). Thus, adaptive management should not be perceived as an end in itself, but rather as a means to galvanize our efforts to advance scientific understanding of the attributes and functioning of an "impaired" environmental system, as well as to crystallize the decision-making process and facilitate the integration of environmental concerns with socio-economic values (Williams and Brown, 2014).

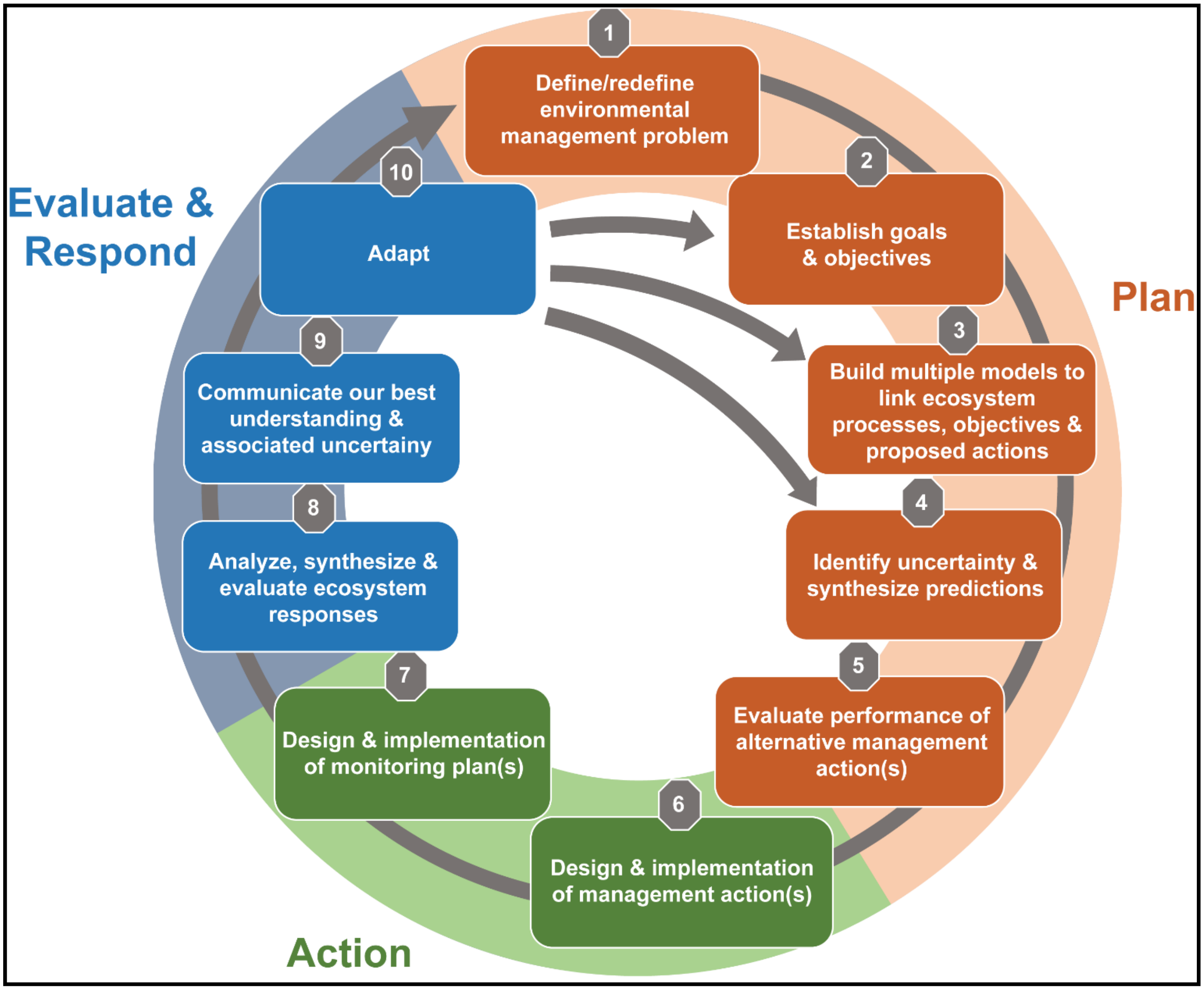


Figure 1: The iterative *monitoring-modelling-assessment* cycles of adaptive management to reduce the environmental uncertainty. In Lake Erie, we currently evaluate steps 4 and 5 in order to design steps 6 and 7.

Environmental modelling is one of the pillars of the adaptive management process, serving as "information integrator" that brings together scientists, managers, and other stakeholders in a joint assessment of the degree of our understanding of the system being managed along with the compelling knowledge gaps we seek to answer through monitoring and research. Models can be qualitative and conceptual, quantitative and detailed, or anywhere in between. As simplistic representations of natural ecosystems, their application in an adaptive management context inevitably introduces the so-called *approximation uncertainty* (Arhonditsis et al., 2018). This uncertainty stems from the imperfect knowledge used and different assumptions made to determine model structure and inputs. Model input error mainly stems from the uncertainty related to the values of model parameters, forcing functions, and initial conditions, as well as from the fact that all models are simplifications of the natural system that approximate the actual processes, i.e., all parameters are effective (e.g., spatially and temporally averaged) values unlikely to be represented by fixed constants (Arhonditsis et al., 2006). Model structure error arises from (*i*) the selection of the appropriate state variables (model endpoints) to reproduce the key physical, chemical, and biological components given the management problem at hand; (*ii*) the selection of the more suitable mathematical expressions among a variety of equations used to describe the same natural process, e.g., linear, quadratic, sigmoidal, and hyperbolic functional forms to reproduce fish predation on zooplankton or Monod vis-à-vis Variable-Internal Stores models to simulate the phytoplankton uptake of nutrients from the water column and their conversion into biomass; and (*iii*) the fact that our models are based on relationships which are derived individually in controlled laboratory environments but may not collectively yield an accurate picture of the entire environmental system (Arhonditsis et al., 2007; 2008a,b). Coupled with other major sources of uncertainty (environmental variation, structural uncertainty, partial controllability, and partial observability), skeptical views question the ability of models to meaningfully assist the decision making process. The counter-argument to this criticism is that the problem of uncertainty is precisely where adaptive management is most valuable (Williams et al., 2009). This paradigm offers a coherent framework to quantify the uncertainty of initial forecasts of management actions and sequentially update them with post-implementation monitoring data; hence, this iterative synthesis of monitoring and modelling provides the basis to reduce the uncertainty and revise management actions (Fig. 2). Even more so, there are voices arguing that adaptive management is better served by introducing multiple competing models that embed alternative hypotheses about the system functioning, as the iterative updating cycles will lend support to the optimal subset of models that more closely reflects the evolution of the environmental system and its potential responses to management actions (Williams et al., 2011; Ramin et al., 2012a).

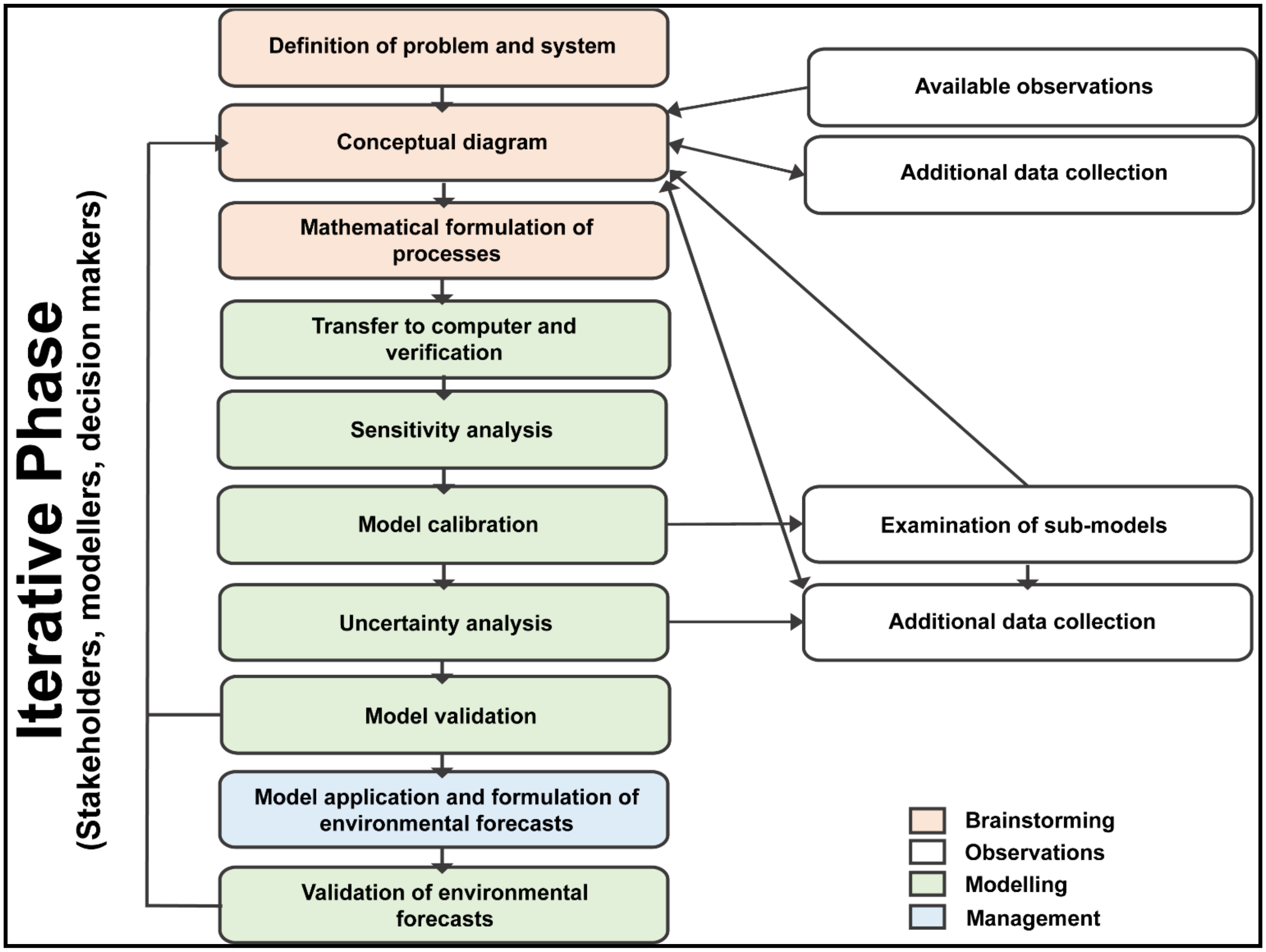


Figure 2: Sequential model updating in the context of adaptive management implementation. The interphase between model uncertainty analysis and data collection is highlighted as a key process to advance our understanding and management actions in Lake Erie.

It is thus axiomatic that the intertwined relationship between monitoring and modelling is a key to the successful implementation of adaptive management. The commitment among managers, scientists, and other stakeholders to establish a sustainable monitoring and assessment program is imperative in this process. Both monitoring and assessment of environmental conditions should be designed to ensure that data are collected for all the relevant resource attributes and within the timeframe required for adaptive-decision making (Williams and Brown, 2014). Another critical aspect of the design of monitoring programs is the appropriate resolution in time and space in order to constrain the existing model(s) and effectively track our progress toward accomplishing management objectives. The information from monitoring is then used to evaluate management, improve understanding, and guide decision making, and it is thus critical to have the appropriate metrics to evaluate the knowledge (or *value of information*) acquired (Yokota and Thompson, 2004; Williams and Johnson, 2015). The latter endeavor represents our *technical learning* of the response of the environmental system to management interventions, which is conceptually (and operationally) nested within the so-called *institutional learning* or the potential to learn about the decision making process itself (Williams and Brown, 2016). Empirical evidence suggests that it may take multiple iterations of the technical learning cycle during which the institutional framework remains practically unaltered, and only a breakthrough in technical learning can prime the pump for revisiting and restructuring the basic institutional elements (Williams and Brown, 2014). More recently, this double-loop learning process has been expanded to distinguish between the socio-political and governance issues typically included within the institutional learning cycle (Pahl-Wostl, 2009; Johnson et al., 2015).

In this study, we present a technical analysis of all the recent modelling work that has been conducted to support the adaptive management process in Lake Erie; the most biologically productive system of the Great Lakes. In response to a binational remedial effort in the late 1970s and early 1980s, west-central phytoplankton biomass and central-basin hypoxia displayed significant improvement in Lake Erie, followed by a general increase since the mid-1990s which was likely the cumulative effect of several stressors, including the invasion of exotic species and increased agricultural loading of bioavailable phosphorus (Watson et al., 2016). Following the evolution of watershed and eutrophication modelling in the literature, Lake Erie represents a unique case study where a wide variety of models have been developed to understand the major causal linkages/ecosystem processes underlying the local water quality problems (Scavia et al. 2016a,b,c). Whether statistical (data-driven) or mathematical (process-based), the basic premise of these models has been to support the decision-making process by linking watershed management actions with the response of the receiving waterbody. We first provide an overview of all the models used in the area and subsequently provide a technical analysis of the forecasts drawn regarding the likelihood to mitigate the eutrophication phenomena in Lake Erie, based on changes in the agricultural management practices followed in the watershed. We then pinpoint knowledge gaps and monitoring assessment objectives that should be addressed to ensure that resource parameters are adequately measured and appropriately focused on relevant performance indicators. Our intent is neither to vilify the modelling enterprise in Lake Erie nor to roundly criticize all the models developed. We recognize that a great deal of modelling work has been done to offer insights into the watershed and lake processes, and our aim with the present study is to impartially identify their strengths and weaknesses. Because of the challenges and complexity of the adaptive framework, we believe that all the experiences gained and lessons learned through the iterative *monitoring-modelling-assessment* cycles will consolidate our know-how with the management of one of the most intensively studied eutrophic systems worldwide (Michalak et al., 2013; Scavia et al., 2017). Our thesis is that critical thinking and effective monitoring are two fundamental prerequisites for hypothesis testing and robust model foundation in order to achieve one of the key aspirations with adaptive management; the reduction of uncertainty.

2 Study Area

Lake Erie is the smallest and shallowest system of the Great Lakes and therefore tends to be the most susceptible to nutrient-driven water quality problems (Fig. 3). The shallowest section of Lake Erie is the western basin where depths average between 7.6 to 9.1 m; the central basin is deep enough to stratify during the summer (mean depth of 18.3 m) typically developing a "thin" hypolimnion with a small volume relative to the epilimnion; the eastern basin has an average depth of 24.4 m and large hypolimnetic volume. Recent evidence suggests that rapid ecological changes have been occurring in Lake Erie that primarily involve the severity of eutrophication phenomena, such as an increase in the magnitude and duration of harmful algal blooms (*HAB*s), prolonged manifestation of hypoxia in the central basin, and excessive *Cladophora* growth in the eastern basin (Higgins et al., 2008; Depew et al., 2011; Michalak et al., 2013; Scavia et al., 2014; Watson et al., 2016). The western part of Lake Erie basin, comprising both the western basin and the Huron-Erie corridor (Detroit River Basin) receives >60% of the external annual total phosphorus (*TP*) loading and among all the tributaries across the entire Lake Erie basin, the Maumee River watershed is the primary contributor of *TP* loading (~30%) into Lake Erie (Maccoux et al., 2016). The catchment area of Maumee River is approximately 16,480 km^2 extending over southern Michigan, northwestern Ohio, and northeastern corner of Indiana (Keitzer et al., 2016). Most of the watershed is predominantly agricultural (73%) and is characterized by a flat landscape (average slope: <2%) and poorly-drained soils. More than 90% of the agricultural land is drained by ditches and subsurface drainage, and 85% of phosphorus exports are reported to originate from agricultural inputs (e.g., fertilizer and manure applications) (Culbertson et al., 2016).

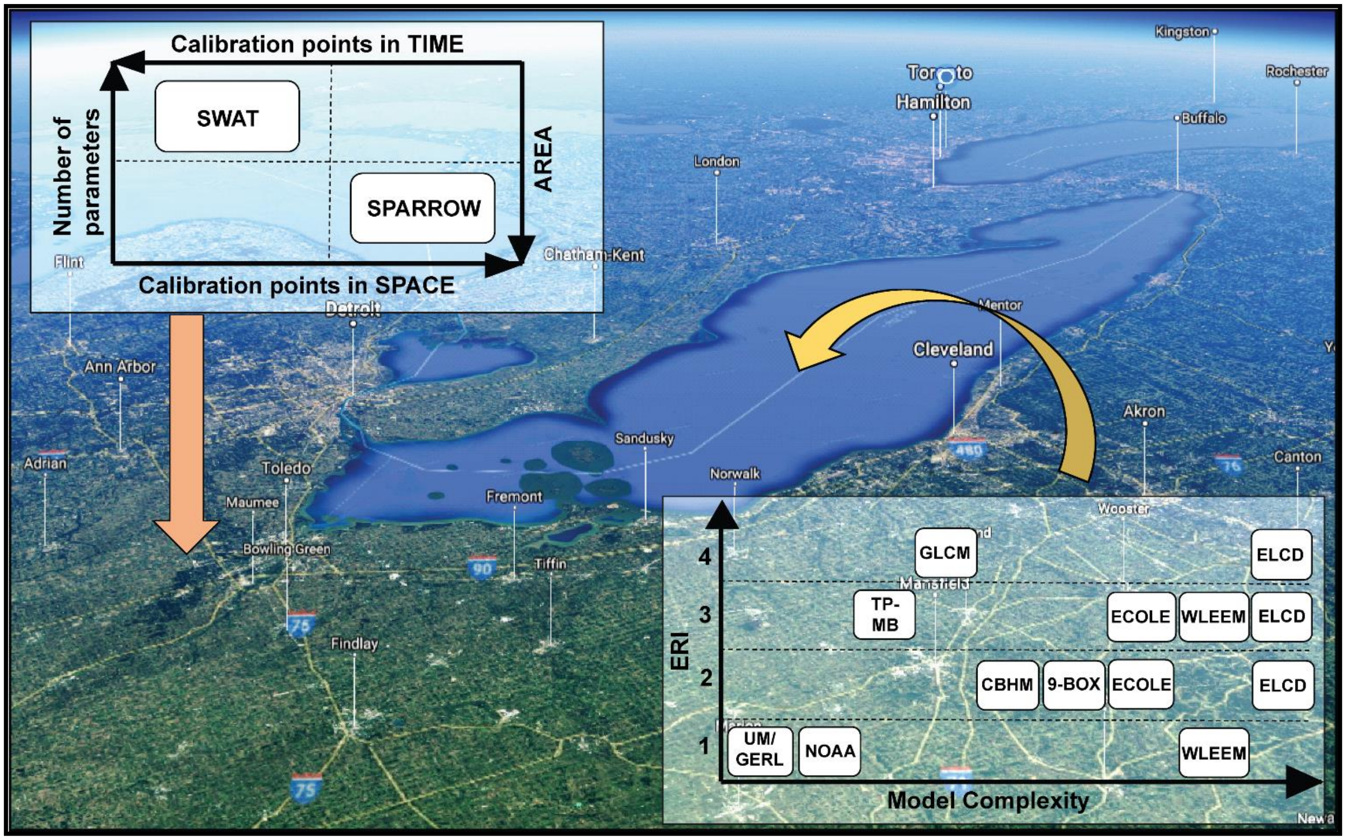


Figure 3: Ensemble modelling in Lake Erie. In the Maumee River watershed, five *SWAT* applications with distinct input assumptions and process characterizations have been used to simulate flow and phosphorus loading at a single downstream station (Waterville, OH), whereas *SPARROW* was used to validate a post-hoc delineation of locations with higher propensity of phosphorus export (top panel). In Lake Erie, a range of data-oriented and process-based models have been used to examine the impact of nutrient loads to ecosystem integrity and to evaluate the achievability of four Ecosystem Response Indicators (*ERI*s) under different watershed management scenarios. Numbers in the Y-axis of the bottom panel stand for: (*1*) western basin cyanobacteria biomass, represented by the maximum 30-day average cyanobacteria biomass; (*2*) central basin hypoxia, represented by number of hypoxic days; average extent of hypoxic area during summer; and average hypolimnion *DO* concentration during August and September; (*3*) basin-specific overall phytoplankton biomass, represented by summer average chlorophyll *a* concentration; (*4*) eastern basin *Cladophora*, represented by dry weight biomass and stored *P* content. The models included in the multi-model ensemble strategy are: *UM/GLERL* Western Basin HAB model; *NOAA* Western Basin *HAB* model; Total Phosphorus Mass Balance Model (*TP-MB*); 1-Dimensional Central Basin Hypoxia Model (*CBHM*); Ecological model of Lake Erie (*ECOLE*); Western Basin Lake Erie Ecosystem Model (*WLEEM*); Estuary and Lake Computer Model-Computational Aquatic Ecosystem Dynamics Model (*ELCD*); Eastern Basin *Cladophora* Model (*GLCM*).

Planktonic blooms in Lake Erie have been characterized by a major shift in the dominant species and particularly a rise in the predominance of toxic cyanobacteria (c*HAB*s), such as *Microcystis* and other potentially toxic (*Dolichospermum, Planktothrix*) species (Millie et al., 2009; Allinger and Reavie, 2013; Kane et al., 2015). The majority of these c*HAB*s frequently occur in the western basin of Lake Erie, where they cover extensive areas and can persist throughout the summer until late fall (Michalak et al., 2013; Steffen et al., 2014). Interestingly, the increased frequency in c*HAB*s has occurred without distinct trend in total annual *TP* loading, but may be related to the timing, sources, and increased bioavailability of *TP* inflows and/or the increased frequency of extreme flow events (Scavia et al., 2014; Obenour et al., 2014). The c*HAB* severity has been causally linked to the volume of spring runoff from the Maumee River watershed, as well as to the elevated dissolved phosphorus inputs from the western basin tributaries (Stumpf et al., 2102; Watson et al., 2016). Growing evidence also highlights the broader role of dreissenid mussels in the ambient conditions (enhanced water clarity, selective particle removal, and soluble nutrient recycling), the supply and chemical speciation of *N*, the increasingly calm summer conditions, and the increasing reservoir of *Microcystis* seed colonies within Lake Erie's benthos (Munawar et al., 2008; Rinta-Kanto et al., 2009; Horst et al., 2014; Davis et al., 2015).

The proliferation of *Cladophora* along the northern shores of the eastern basin since the mid-1990s, was primarily attributed to the increased water clarity and suitable substrate following the colonization of the area by dreissenid mussels (Higgins et al., 2005). Nutrient-rich hypolimnetic masses of water transported in nearshore through upwelling events, excreted metabolic wastes, and/or particulate matter available for re-mineralization through egestion of non-edible algae by dreissenids are the likely suppliers of nutrients in the benthic environment (Wilson et al., 2006; Depew et al., 2011; Valipour et al., 2016). The magnitude, duration, and frequency of hypoxia is predicted to be exacerbated by climate change via a number of mechanisms such as increased water temperature, deeper and longer stratification, and increased nutrient runoff during winter and spring (Fang and Stefan, 2009; Jiang et al., 2012). Because of its bathymetry, the severity of hypoxia in central Lake Erie can display both intra- and inter-annual variability driven by the impact of local weather conditions (e.g., wind, temperature) on physical processes such as mixing, within- and between-basin circulation, and water-column stratification (strength, depth), which in turn influence the rate of dissolved oxygen (*DO*) depletion, sediment-oxygen demand, and *DO* transfer across the thermocline (Zhou et al., 2013; Rucinski et al., 2014). Consistent with these predictions, a record-breaking hypoxic event occurred in Lake Erie in 2012, following a period of prolonged drought and low tributary flows (Stow et al., 2015).

3 Model Ensembles and Adaptive Management Implementation

Recognizing the need to address these ominous water quality trends, the International Joint Commission (*IJC*) formulated the Lake Erie Ecosystem Priority taskforce in 2012 to develop a sustained restoration plan by identifying knowledge and monitoring gaps, evaluating current conditions, and providing guidance for management targets (*IJC*, 2014; Watson et al., 2016). This initial *IJC* review, and associated work by binational taskforces, led to commitments for remedial action in the 2012 renewed Canada-*USA* Great Lakes Water Quality Agreement (*GLWQA*). Nutrients, algal biomass, and hypoxia are addressed under Annex 4 of this Agreement, with special emphasis on setting interim *TP* load and basin-specific concentration targets for Lake Erie. As part of the *GLWQA* review, a committee of modellers evaluated a set of Great Lakes eutrophication models that were designed to establish target *TP* loads in order to mitigate the eutrophication symptoms in the early 1970s (Scavia et al., 2016a). While the post-audit performance of the original eutrophication models in Lake Erie was satisfactory (e.g., Di Toro et al., 1987; Lesht et al., 1991), their basic structure was not deemed adequate to reproduce the relative importance of nearshore processes, the wide array of factors triggering *HAB* formation or the proliferation of nuisance benthic algae, and the role of other external stressors, such as invasive species or climate change (Scavia et al., 2016a,b,c). To address the wide range of conceptual and operational uncertainties typically characterizing any modelling exercise, the local community opted for a novel multi-model strategy that aimed to capitalize upon the wide variety of both empirical and process-based models of variant complexity that have been developed for Lake Erie over the past decade (Scavia et al., 2016a,b,c).

Being primarily a reflection of our current level of understanding and existing measurement technologies, the multi-model strategy adopted for Lake Erie accommodates the fact that many different model structures and many different parameter sets within a chosen model structure can acceptably reproduce the observed behavior of a complex environmental system (Beven and Freer, 2001; Christakos, 2003; Arhonditsis et al., 2007; 2011). While this very important notion is still neglected in the modelling literature, there are viewpoints suggesting that environmental management decisions relying upon a single, *partially adequate* model can introduce bias and uncertainty that is much larger than the error stemming from a *partially defensible* selection of model parameter values (Neuman, 2003). Importantly, the practise of basing ecological predictions on one single mathematical model implies that a valid alternative model may be omitted from the decision making process (Type I model error), but also that our forecasts could be derived from an erroneous model that was not rejected in an earlier stage (Type II model error). Recognizing that there is no true model of an ecological system, but rather several adequate descriptions of different conceptual basis and structure, the development of model ensembles is a technique specifically designed to address the uncertainty inherent in the model selection process. Instead of picking the single "best-fit" model to draw ecological forecasts, we can use a multi-model ensemble to derive a weighted average of the predictions from different models (Ramin et al., 2012a).

Notwithstanding the voices in the literature asserting that we are still missing rigorous methodological frameworks to develop multi-model ensembles (Neuman, 2003), the basic framework comprises several steps related to the development of "truly" distinct, site-specific conceptual models, selection of the optimal subset of both data-driven and process-based models, effective combination of these models to synthesize predictions, and subsequent assessment of the underlying uncertainty (Fig. 4). This methodological procedure involves three critical decisions aiming: (*i*) to identify the conceptual or structural differences of the existing models (*ensemble members*), and thus determine the actual diversity collectively characterizing the model ensemble; (*ii*) to determine the most suitable calibration/validation domain for evaluating model performance in time and space; and (*iii*) to establish an optimal weighting scheme in order to assign weights to each ensemble member, when integrating over the individual predictions, and determine the most likely outcome along with the associated uncertainty bounds (Ramin et al., 2012a). In this study, we dissect the two model ensembles developed for the Maumee River watershed and the Lake Erie itself and evaluate their compliance with the aforementioned framework (Fig. 3).

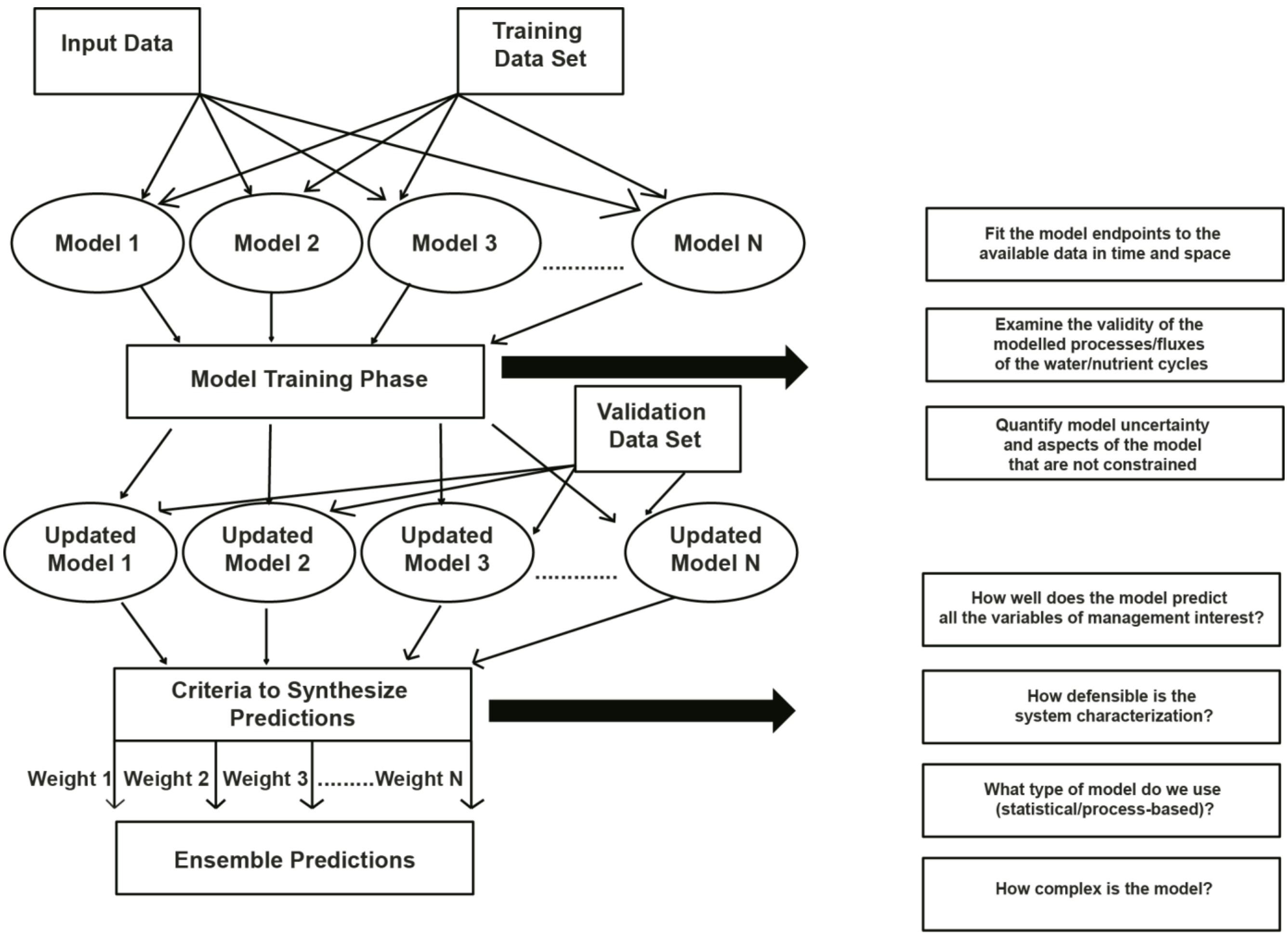


Figure 4: Ensemble modelling is the process of running two or more related (but different) models with respect to their conceptual/structural characterization and input specification, and then synthesizing the results into a single score or spread in order to improve the accuracy of predictive analytics and data mining applications. The development of a weighting scheme to determine the relative contribution of each model to the ensemble predictions is missing from the current Lake Erie modelling work

In terms of model diversity, the watershed modelling work has been based on five applications of the same complex mathematical model (i.e., *SWAT*), each application reflected an independent attempt to characterize the processes that modulate phosphorus export from Maumee River. As such, even though the mathematics were the same (i.e., *SWAT* equations), the five models collectively capture some of the uncertainty related to our understanding of watershed attributes and functioning, and thus this exercise qualifies to be treated as an ensemble strategy (Kim et al., 2018a). Alongside the five *SWAT* applications, an empirical model (*SPARROW*) was also used to validate a post-hoc delineation exercise of locations with higher likelihood of phosphorus export in the Maumee River watershed. On the other hand, the lake ensemble has been based on a wide range of data-driven and process-based models to examine the achievability of different environmental targets under various external nutrient loading regimes (Shimoda et al., 2018). Beyond the structural diversity of the models in place, our study also attempts to identify their actual variability in terms of the characterization of fundamental biogeochemical processes modelled, ecological insights gained, and the nature of predictions drawn. Our thesis is that the true diversity of a model ensemble is not solely determined from the complexity of the mathematics or the number of system components simulated, but also from the specification of the ecosystem functioning, as determined by the decisions made during the calibration and validation phase of the individual models, which ultimately shape the forecasts derived to guide the policy-making process. Regarding the model assessment, the typical criteria used are the degree of prediction errors (*skill*; the difference between predictions and observations); the relative goodness-of-fit for the different model endpoints (*bias*; differences in the model fit among flow, total and dissolved reactive phosphorus loading along with the nature of the error, i.e., over- or under-estimation); and the degree of divergence of the individual model forecasts when forced with various management scenarios (*forecasting spread*) (Arhonditsis and Brett, 2004; Arhonditsis et al., 2011). In this study, we will examine if the three criteria have been consistently evaluated, to what extent the model assessment was based on suitable calibration/validation domains and spatiotemporal scales relevant to the environmental management problems in question, and whether or not the performance assessment was used to determine the relative weights that each model carried during the synthesis of the individual predictions.

3.1 *SWAT*-based Ensemble Strategy for the Maumee River Watershed: Description, Performance, and Outstanding Questions.

In the Great Lakes area, a wide range of process-based models have been used to characterize the watershed processes associated with the hydrological cycle and nutrient fate and transport, but the Soil and Water Assessment Tool (*SWAT*) model alone represents more than >70% of the watershed modelling studies published for Lake Erie in the peer-reviewed literature (Kim et al., 2018a; Dong et al., 2018). *SWAT* is a physically-based, semi-distributed watershed model that can operate on various time (sub-daily to annual) steps and its structure can accommodate the weather generation, watershed topography, hydrological processes, and agricultural practices (Neitch et al., 2011). The watershed is disaggregated into subbasins and these subbasins are subsequently subdivided into hydrological response units (*HRUs*), which represent distinct combinations of soil and land use characteristics. The water balance of each *HRU* in the watershed is based on four storage volumes: snow, top soil (0-2 m), shallow aquifer (2-20 m), and deep aquifer (>20 m). Surface runoff is commonly estimated using the Curve Number (*CN*) method, and sediment erosion is calculated with the Modified Universal Soil Loss Equation (Arnold et al., 1998). Nutrient (nitrogen and phosphorus) dynamics are reproduced by considering both mineral (stable, active, and soluble) and organic (stable, active, and fresh) forms. The preference for *SWAT* over other watershed models is likely due to its ability to carry continuous long-term simulations in predominantly agricultural watersheds, as well as to reproduce the impact of episodic rainfall events at finer (i.e., daily or sub-daily) resolutions (Neitch et al., 2011; Wellen et al., 2014a,b). While *SWAT* has a well-documented potential to offer an insightful tool for the examination of agricultural best management practices (*BMP*s), it should be noted that there are other models (Storm Water Management Model or *SWMM*, *MIKE SHE*) available with technical features more relevant for the representation of urban environments (Dong et al., 2018; see also following discussion).

Like many watershed models, *SWAT* is basing its predictive capacity on the explicit consideration of a wide array of physical, chemical, and biological processes that can shape downstream flow and nutrient export conditions. This strategy -in principle- renders greater assurance that all the potentially important mechanisms operating in a watershed context are accounted for, and thus the application domain of the model can be effectively extended (Beven, 2006; Pappenberger and Beven, 2006). However, increasing model complexity inevitably magnifies the disparity between what we want to learn from a model and what we can realistically measure/monitor given the available technology or resources at hand (Arhonditsis et al., 2008a,b; Rode et al., 2016). Opting for an increased model complexity without a commensurate increase in the available empirical knowledge exacerbates the problem of under-determination, and thus we experience a situation whereby several distinct choices of model inputs lead to the same model output, or alternatively many sets of parameters fit the data about equally well. This non-uniqueness of model solutions, also known as equifinality, negates the main function of mathematical models as inverse analysis tools, i.e., the available data for the dependent variables (flow, suspended solids, nutrient concentrations) are used through the model training (or calibration) phase to advance our knowledge of independent variables (model parameters) typically representing fundamental processes/fluxes of the water budget and/or nutrient cycles. To overcome the latter problem, the modelling work in Maumee River watershed was built upon five *SWAT* model applications from groups affiliated with Heidelberg University (*HU*), LimnoTech (*LT*), Ohio State University (*OSU*), Texas A&M University (*TAMU*), and the University of Michigan (*UM*) (Culbertson et al., 2016; Gildow et al., 2016; Kalcic et al., 2016; Keitzer et al., 2016; Muenich et al., 2016). Based on different assumptions and independent calibrations, each model reflects a distinct characterization of the watershed processes modulating flow dynamics and phosphorus loading export. Thus, the five *SWAT* applications collectively represent an attempt to capture our incomplete knowledge of watershed attributes and functioning, and create an uncertainty envelope that can be propagated when the models are used for forecasting purposes and land-use management scenarios (Kim et al., 2018a; Dong et al., 2018). It is also important to note that the characterization of the processes associated with the nitrogen cycle has not been studied yet, even though the relative contribution of different nitrogen (*N*) species from the Maumee River watershed is likely to be one of the regulatory factors of the downstream water quality conditions; especially the composition of the phytoplankton community.

The calibration of the five *SWAT* models covered different the time spans, 4 years for *HU* (2009-2012), 13 years for *LT* (1998-2013), 10 years for *OSU* (2000-2009) and *TAMU* (1990-1999), and 5 years for *UM* (2001-2005) (Scavia et al., 2016c). Consequently, each model used different landscape (e.g., land-use/land-cover data) and meteorological information (see Table 1 in Kim et al., 2018a). Moreover, the five *SWAT* models differed with respect to the agricultural inputs used, such as fertilizer and manure application rates. In particular, each model postulated different fertilizer application rates from different sources of information, such as tri-state standards for *HU* and *OSU*, *USDA-ARS* for *LT*, agriculture census data for *TAMU*, and county fertilizer sales data for *UM*. Regarding the manure application, only two models (*LT* and *UM*) explicitly considered the role of this potentially important source of organic nutrient release in the soils. Another critical source of uncertainty revolved around the tile drainage assumptions. Three of the *SWAT* applications (*LT*, *OSU*, and *UM*) postulated that tiles are deployed at poorly-drained agricultural lands, but each model was based on different definitions of the categories of poorly-drained conditions (see Table 1 in Kim et al., 2018a). Two models, *HU* and *TAMU*, also assumed that the tile drainage of the Maumee River watershed is highly correlated with low-slope areas, i.e., the *HU* model used a 3% slope to distinguish between steep and flat areas, while the cutoff point for *TAMU* was less than 1% slope. All five *SWAT* models were calibrated against a single downstream station (Waterville, *OH*), while the *LT* model used two additional upstream sites (Blanchard and Tiffin) for a subsequent validation exercise (Scavia et al., 2016c). Consistent with the methodological practice recommended in the literature (Fig. 4), the five *SWAT* applications were based on identical validation time periods (from 2005 to 2014), which in turn (partly) insured their comparability for the purposes of a post-hoc ensemble synthesis (Kim et al., 2018a).

Table 1. Predictive validation of the five SWAT models against monthly flow and phosphorus loading at Waterville, Ohio, from 2005-2014. Values close to 0 for percent bias (PBIAS) and 1 for the Nash-Sutcliffe efficiency (NSE) and the coefficient of determination (r^2) indicate higher model performance.

	Measures of Model Fit*	*Models*					
		HU	*LT*	*OSU*	*TAMU*	*UM*	Model Average
Flow	*PBIAS*	-7%	10%	10%	11%	6%	6%
	MEF	0.82	0.90	0.91	0.86	0.89	0.88
	r^2	0.86	0.91	0.93	0.88	0.91	0.90
TP	*PBIAS*	37%	-6%	-7%	-22%	7%	2%
	MEF	0.64	0.82	0.73	0.56	0.70	0.69
	R^2	0.74	0.82	0.75	0.71	0.70	0.75
DRP	*PBIAS*	81%	1%	16%	-13%	-13%	14%
	MEF	-0.02	0.71	0.51	0.52	0.46	0.44
	r^2	0.55	0.71	0.54	0.70	0.51	0.60

$* \ PBIAS = \frac{\sum(Observed - Predicted)\times 100}{\sum Observed}$ $MEF = 1 - \frac{\sum(Observed - Predicted)^2}{\sum(Observed - Observed\ Average)^2}$

$r^2 = \frac{\sum[(Observed - Observed\ Average)\times(Predicted - Predicted\ Average)]}{\sum(Observed - Observed\ Average)^2 \times \sum(Predicted - Predicted\ Average)^2}$

Prior to conducting the validation exercise, Scavia et al. (2016c) specified the criteria of excellent fit against the flow rates to be lower than ±10%, higher than 0.5, and higher than >0.6 for the percent bias (*PBIAS*), modelling efficiency (*MEF*), and coefficient of determination (r^2), respectively. Likewise, the excellent performance for phosphorus loading was determined to be *PBIAS* <±25%, *MEF* >0.4 and r^2 >0.5. Based on these pre-specified performance standards, the five *SWAT* models showed nearly excellent goodness-of-fit against measured monthly flow rates and phosphorus loading empirical estimates (Table 1). The *OSU* model reported the best agreement against the measured flow rates in terms of *MEF* (0.91), *PBIAS* (10%), and r^2 (0.93). By contrast, the *HU* model was the only *SWAT* model that underestimated the flow rates (*PBIAS*: -7%) and showed the lowest *MEF* (0.82) and r^2 (0.86) among the five models. Regarding the *TP* and dissolved reactive phosphorus (*DRP*) loading predictions, the individual *SWAT* applications presented lower goodness-of-fit for both estimates, but the performance for *TP* (*MEF*: 0.56-0.82 and r^2: 0.70-0.82) was higher than for *DRP* (*MEF*: -0.02-0.71 and r^2: 0.51-0.71). The *LT* model outperformed the other four *SWAT* applications in its ability to capture the monthly *TP* and *DRP* load variability, whereas the *HU* model significantly overestimated the *TP* (*PBIAS*: 37%) and *DRP* (*PBIAS*: 81%) loads. Given that the same *SWAT* application underestimated the flow rates, the latter result suggests that the simulated concentrations were distinctly overestimated to the extent that not only compensated for the underpredicted flows but also led to excessively higher phosphorus loading predictions (Kim et al., 2018a). By contrast, the *TAMU* model clearly underestimated *TP* and *DRP* loads (-22% and -13%, respectively) comparing with the overestimated flow rates (11%), which again is indicative of an even greater bias (i.e., underprediction) for the corresponding concentrations.

One of the critical skill assessment tests of a dynamic watershed model, like *SWAT*, is its ability to support predictions at a finer (daily or subdaily) time scale (Wellen et al., 2015. Surprisingly, the *SWAT* modelling work in the Maumee River watershed placed little emphasis on evaluating the robustness of the hydrological or nutrient loading predictions with such a temporal resolution. To shed light on this important issue, Kim et al. (2018a) independently evaluated the error associated with the daily outputs from the *UM* model and showed that the simulated *TP* and *DRP* concentrations are characterized by a substantial error (see their Figs 4 and S2). Importantly, the same study evaluated the ability to capture the impact of episodic precipitation events, by using a peak-flow threshold of 1000 $m^3\ s^{-1}$ to separate extreme events from the rest of the flow conditions over the course of a 5-yr period (2001-2005). The results showed that the *UM SWAT* consistently underestimated the flow rates in 20 out of 22 events (≈ 91%), reinforcing the point that an excellent goodness-of-fit with a coarser (seasonal or monthly) resolution may simply stem from multiple daily errors/biases that cancel each other out when seasonally or monthly averaged, and thus does not necessarily guarantee acceptable performance against finer time scales (Arhonditsis and Brett, 2004). Given that the *UM* displayed the lowest discrepancy against the monthly flows (*PBIAS*: 6%) and *TP* loading (7%) among the five *SWAT* applications (Table 1), it stands to reason that the rest of the models likely fare worse with their daily simulations. If we also consider the emerging evidence of a disproportionally high fraction of the total annual loading occurring during event-flow conditions, it is particularly critical to strengthen the ability of the *SWAT*-ensemble to assess the impact of episodic/extreme rainfall events and elucidate their role on the broader watershed dynamics.

The potential impacts of climate change on hydrological extremes have received considerable attention during the last decade. It is predicted that global warming will amplify the hydrological cycle, and thus less frequent but more intense precipitation events could increase the severity of within-season drought, alter evapotranspiration, and generate greater runoff (Huntington, 2006; Knapp et al., 2008; Goodess, 2013). In the same context, recent empirical evidence is on par with these projections suggesting that (*i*) total phosphorus and phosphate loads can vary by three orders of magnitude between wet and dry conditions; (*ii*) storm events and spring freshets play a predominant role with the peak daily loads in urban and agricultural watersheds, respectively; (*iii*) a significant fraction (>50%) of the annual phosphorus loads can be generated during a small number of brief but intense precipitation events; and (*iv*) the flow-concentration relationship can be significantly influenced by the watershed physiography, land-use patterns, and antecedent soil moisture conditions (Green et al., 2007; Long et al., 2014; 2015). On the other hand, an overarching flow-concentration paradigm for *N* species is even less clear relative to that for phosphorus. Contrary to our understanding of *TP* fate and transport, a greater proportion of total nitrogen (*TN*) is found in the dissolved phase due to relatively high solubility of nitrogen species, such as nitrite and nitrate, and can be transported by both overland and subsurface flow paths (Long et al., 2014). Furthermore, subsurface leaching of nitrate, and hence transport to groundwater, is generally greater than phosphate due to immobilization of phosphate by clay and other soil chemical constituents. Thus, one of the working hypothesis is that phosphorus and certain *N* species (ammonia, total Kjeldahl nitrogen) will be distinctly higher during precipitation and snowmelt events, while nitrate will display chemostatic behavior (i.e., apparent stability of the concentrations relative to the flow variability) (Godsey et al., 2009; Long et al., 2014). Another testable hypothesis is that the nutrient loads per unit area will be significantly higher from urban relative to agricultural watersheds and may be further exacerbated with climate change, as hydrological behavior can change above certain discharge or precipitation thresholds (Wellen et al., 2014a,b; Kim et al., 2016, 2017). Climate warming will increase the vulnerability of soils to erosion in winter (snowpack decrease, early onset of spring snowmelt, frequent rainfall events, and snowmelt episodes), and consequently the contemporaneous sediment and nutrient loadings relative to current levels. It is critical to understand the potential changes in the interplay among catchment state, land use, and nutrient export patterns induced by a changing climate, and this is another area where little work has been done in the Maumee River watershed.

3.2 Multi-model Ensemble Strategy for Lake Erie: Description and Performance

Counter to the watershed modelling framework for the Maumee River Watershed, the multi-model approach for the lake itself comprised a range of data-oriented and process-based models to examine the impact of nutrient loads to ecosystem integrity in Lake Erie and to evaluate the achievability of Ecosystem Response Indicators (*ERI*s) under different watershed management scenarios (Fig. 3; see also Scavia et al., 2016a,b). The primary advantage of the suite of models used was their ability to capitalize upon the advantageous features of both statistical and process-based approaches; namely, the former models are derived from empirical parameter estimation that allows for rigorous assessment of predictive uncertainty, while the latter ones have the mechanistic foundation that can conceivably enable to draw predictions outside the domain used during their calibration or even validation (Arhonditsis et al., 2007). Regarding the diversity of the process-based modelling work for Lake Erie, the models developed come from the entire complexity spectrum with different strengths and weaknesses (Scavia et al., 2016a,b). There are simple models in place with fewer unconstrained parameters that can be more easily subjected to uncertainty analysis, but they are also criticized as being crude oversimplifications not capable of reproducing the wide range of ecosystem behaviours. By contrast, there are complex models with numerous ecological processes and more sophisticated parameterization that are presumably more suitable to capture any potential non-linear system responses to distinct changes in external or internal conditions (e.g., land-use management, climate change, invasive species), but the main criticisms are their inevitably poor identifiability that may lead to ecosystem mis-conceptualizations, as well as their high computational demands that can be prohibitive in comprehensively quantifying their structural or parametric uncertainty (Arhonditsis and Brett, 2004; Arhonditsis et al., 2007).

Specifically, the models included in the multi-model ensemble strategy were two empirical models, *UM/GLERL* Western Basin HAB model (Bertani et al., 2016) and *NOAA* Western Basin *HAB* model (Stumpf et al., 2016), and six process-based models: Total Phosphorus Mass Balance Model (*TPMB*; Chapra et al., 2016), 1-Dimensional Central Basin Hypoxia Model (*1D-CBH*; Rucinski et al., 2016), Ecological model of Lake Erie (*EcoLE*; Zhang et al., 2016), Western Basin Lake Erie Ecosystem Model (*WLEEM*; Verhamme et., 2016), the Estuary and Lake Computer Model-Computational Aquatic Ecosystem Dynamics Model (*ELCOM-CAEDYM* or *ELCD* in Fig. 3; Bocaniov et al., 2016); Eastern Basin *Cladophora* Model (*EBC* or *GLCM* in Fig. 3; Valipour et al., 2016). The multi-model ensemble exercise for Lake Erie also included the Nine-Box model (Lam et al., 2008); a coarse grid (9-box) phosphorus mass balance model designed to offer quantitative understanding of Lake Erie eutrophication (organic and dissolved-phase phosphorus) and related hypoxia (Lam et al., 1983). Although the model was extensively calibrated and validated against measurements from the mid-1960s until the early 1980s, and subsequently recalibrated to reproduce the water quality conditions during the post-dreissenid period, its use for the construction of load-response curves was fairly minimal (Scavia et al., 2016b). It was thus omitted from the present synthesis, but it must be noted that the Nine-Box model has the potential to offer a parsimonious management tool after the addition of several critical variables and ecological processes related to the planktonic food web of Lake Erie (Lam et al., 2008).

In reviewing the pertinent literature, a first notable finding is that the local modelling work has closely followed the recommended methodological protocol when developing models intended to assist environmental management (Fig. 2). Following the evolution of each model over time, we can find detailed sensitivity analysis exercises and goodness-of-fit statistics against a wide range of multi-year conditions, while Scavia et al. (2016a,b) also attempted to draw parallels among the different models and offer a *much-needed* synthesis in order to establish the multi-model forecasting tool. On the other hand, because of the complexity of the existing mechanistic models in Lake Erie, the rigorous quantification of their uncertainty can be particularly challenging and thus has not received significant attention. A plausible way to control model uncertainty is to capitalize upon the findings of the sensitivity analysis exercises presented in the local literature by obtaining empirical estimates for critical/influential parameters or other inputs, which can be directly measured in the field or experimentally quantified (see also discussion in Kim et al., 2018a). There are also promising statistical ways (e.g., linear or non-linear emulators) to overcome the cumbersome structure of complex models (e.g., *ELCOM-CAEDYM*, *WLEEM*, *EcoLE*) and quantify their uncertainty that should be considered in the next iteration of the modelling framework (Kim et al., 2014). Consistent with the general trend in the international modelling literature (Arhonditsis and Brett, 2004), we found that the performance of the aquatic ecological models in Lake Erie declined from physical-chemical to biological components of planktonic systems (Fig. 5). Regarding our skill assessment analysis, it is important to note that it is based on a point comparison in time and space, as oppose to the aggregated spatiotemporal (basin- or lake-wide, seasonal/annual time) scales adopted from the local modelling community (see following discussion). Counter to the *SWAT*-ensemble framework, it is also worth mentioning that there is no identical validation time period, within which the performance of all the process-based models across different water-quality variables has been examined, and thus –strictly speaking- the basic condition to ensure their comparability for the purposes of a post-hoc ensemble synthesis is not met. A brief description of the basic technical features and performance against existing data of each member of the multi-model ensemble is provided in the following section, while more details about their structure and assumptions can be found in Scavia et al. (2016a,b) and references therein.

U-M/GLERL Western Lake Erie HAB model: This is a regression model originally developed by Obenour et al. (2014) to predict peak *HAB* size in western Lake Erie as a function of spring *TP* loading from the Maumee River. An interesting feature of this empirical construct is the consideration of a temporal trend component aiming to capture the variability over time in the lake's susceptibility to *HAB* formation. The optimal *TP* loading period (i.e., January-June) for predicting bloom size was assessed probabilistically using a weighting parameter that represents the temporal threshold, prior to which loading does not contribute to the late-summer algal bloom. More recently, Bertani et al. (2016) further refined the model by introducing an expression to estimate probabilistically the fraction of the particulate *P* load that becomes bioavailable. The model was jointly calibrated against three sets of bloom observations: (*i*) *MERIS* satellite remote sensing imagery (Stumpf et al., 2012); (*ii*) in situ measurements from a plankton sampling program (Bridgeman et al., 2013); and (*iii*) *SeaWiFS* satellite remote sensing imagery (Shuchman et al., 2006). A Bayesian hierarchical configuration enabled the characterization of the observation error with each dataset as well as the realistic quantification of the total predictive uncertainty of the derived annual bloom size. In its most recent calibration, the model accommodated the impact of dreissenids on *P* recycling by increasing the bioavailability of particulate phosphorus (*PP*) to algae (i.e., a prior probability was assigned to the relevant parameter postulating that the readily bioavailable *PP* discharged from tributaries should be at least 20%). The model explained over 91% of the year-to-year variability in bloom observations over the course of seventeen (17) years (1998-2014), while cross-validation performance remained relatively high, 80.6% (Scavia et al., 2016b).

These results are certainly promising, but it is important to note that the sample size (or time span covered) for the model training phase is still small, and thus each annual observation (i.e., pair of spring loading and bloom size values) carries significant weight on the form of the resulting equation as well as on the predictions drawn. In particular, a leave-one-year-out cross-validation exercise (i.e., removal of observations from each year, re-calibration of the model against the remaining data, and then forecasting of the excluded observations) revealed wide ranges for all the posterior estimates of the model parameters (see Table B2-4 in Scavia et al., 2016b). Along the same line of evidence, Bertani et al. (2016) noted that the additional years used for the calibration contain the markedly higher bloom sizes in 2013 and 2014 than in previous years, suggesting that the lake may becoming more susceptible to cyanobacteria blooms over time, with smaller loads apparently triggering larger blooms in most recent years. The latter finding was on par with Obenour et al.'s (2014) positive posterior estimate of the temporal trend term, which reinforces the point of an increased propensity for bloom formation in Lake Erie over time. A characteristic example of the latter finding is that the model predicts a bloom size of 15,700 MT (95% predictive interval: 5,000-23,700) associated with a spring weighted *TP* load of 204 MT/month for the 2013 conditions, while the same bloom size prediction is provided by twice as high weighted TP load in 2008 (Scavia et al., 2016b).

NOAA Western Lake Erie HAB model: This is another regression model that attempts to elucidate the primary factors modulating the severity of summer cyanobacterial blooms in western Lake Erie, which in turn was first quantified by the Medium Resolution Imaging Spectrometer (*MERIS*) satellite imagery data. Using as a surrogate of total cyanobacteria biomass, the "cyanobacteria index" (*CI*) (Wynne et al., 2008; 2010), the original models presented by Stumpf et al. (2012) were based on a 10-year dataset (2002-2011). With the loss of *MERIS* data in April 2012, the Moderate Resolution Imaging Spectroradiometer (*MODIS*) data were employed for the subsequent four years (Stumpf et al., 2016), but it was noted that *MERIS* produces a more sensitive proxy (nominal uncertainty 10%) with less noise than *MODIS* (nominal uncertainty 25%). The *CI* is determined in 10-day composites by taking the highest cyanobacterial chlorophyll-related index at each pixel available from any of the daily images within a given 10-day time period, and thus remove interference and biasing from clouds. The annual *CI* used to define the bloom severity is the average of the three 10-day periods around the maximum severity of the bloom, so it is effectively a 30-day average which is then converted to biomass using a ratio of 4800 MT per *CI* (Scavia et al., 2016b).

According to Stumpf et al. (2016), the Maumee River discharge and total bioavailable phosphorus (*TBP*) loading from March through July, with July excluded only when June water temperatures were below the optimal temperature (20°C), were the best predictors for *Microcystis* growth. Drawing similarities with Bertani et al. (2016), *TBP* was specified as the sum of the dissolved reactive phosphorus and the proportion of particulate phosphorus that is bioavailable, corrected for loss due to settling in the river. For the purposes of the multi-model ensemble forecast exercise, Scavia et al. (2016b) developed a variation of the previous models using as predictor the *TP* loading from March to July for all years, except 2003 and 2008, when the July temperature average was lower than 18°C. The latter model explained 62% on the observed variability of the annual cyanobacteria bloom *CI* data, but the standard error of the slope was 24% of the mean slope estimate, and the uncertainty factor was approximately equal to 1.95 (or simply put, the *CI* predictions derived by the model are correct within a factor of 2). Similar to the *U-M/GLERL* Western Lake Erie *HAB* model, the "leave-one-out" test was indicative of a substantial sensitivity of the model intercept and slope to the individual years included in the calibration dataset; especially when the *CI*-*TP* loading pair for 2012 is considered (see Table B1-1 in Scavia et al., 2016b). Considering that this is a simple exponential model (i.e., without the ability to accommodate the long-term trends in the susceptibility of Lake Erie), the analysis of *TP* loading scenarios provided very optimistic forecasts, suggesting that a 40% load reduction will significantly reduce the likelihood of severe cyanobacteria blooms, even for years when river discharge will be excessively high (Scavia et al., 2016b). To put these forecasts into perspective, Stumpf et al. (2016) cautioned that the model in its current exponential form may not be able to capture the biomass-loading relationship during extreme years, like the 2015 bloom, and instead a logistic (sigmoidal-type) function could be a more suitable model. Notwithstanding the plausible ecological arguments to adopt an equation that postulates a "saturation-type" pattern (or a smaller increase rate of the bloom size) when excessively *TP* loading are experienced in Lake Erie (Stumpf et al., 2016), an alternative explanation could be that the capacity of satellite images to characterize the algal bloom severity has an upper bound, as the satellite data cannot quantify the additional biomass production once scum completely covers the entire area of water observed in each pixel.

Total Phosphorus Mass Balance Model: This is a simple phosphorus budget model originally used to establish phosphorus loading targets for the 1978 Great Lakes Water Quality Agreement (Chapra, 1977; Chapra and Robertson, 1977; Chapra and Sonzogni, 1979). The governing model equation predicts offshore *TP* concentrations, as a function of external loading, inter-segment hydrodynamic exchanges, and net sedimentation losses. Evolved from the original spatial segmentation, the model currently divides Lake Erie into three completely mixed compartments (western, central, and eastern basins), while the within-lake intersegment flow exchanges were derived by annual water balance estimates (tributary flows, lake level variability, and over-lake precipitation minus evaporation). Alongside the advective transport, the model considers a bulk mixing coefficient which is a phenomenological parameter aiming to capture large-scale diffusive exchanges across open boundaries, such as large-scale eddy diffusion, and dispersion due to shear flow and spatial non-uniformities. The latter feature was parameterized with a chloride budget model (see Table 3 in Chapra et al., 2016), and the *TPMB* model was subsequently calibrated against median offshore *TP* concentrations by simply adjusting the segment-specific, apparent settling velocities. Thus, the model offers a parsimonious construct to obtain first-order approximations of the in-lake processes that collectively modulate the ambient *TP* variability. The goodness-of-fit statistics were suggestive of satisfactory *TPMB* performance with a root mean square error between 3.6-5.4 μg L^{-1} and a percent relative error (*RE*) between 26-29% (study period 1970-2010). The discrepancy between modelled and measured *TP* concentrations was attributed to the predominant role of the basin-specific tributary loadings, but other factors, such as errors stemming from the intersegment advective/diffusive mass exchanges or the segment-specific apparent settling velocities, were not ruled out (Chapra et al., 2016).

According to Occam's razor (or principle of parsimony), models should be as simple as possible, but not simpler to the extent that we fail to consider important facets of the system modelled (Paudel and Jawitz, 2012). The *TPMB* application in Lake Erie is no exception, as the fitting exercise pinpointed the need to assign higher apparent settling velocities in order to reproduce the higher *P* retention during the post-1990 period. A suite of mechanisms were suggested to provide the ecological underpinning of such structural augmentation, such as the dreissenid filtration that enhanced transport to the sediments; the proliferation of soft sediment settlers, i.e., quagga mussels (*Dreissena bugensis*), that may contribute to permanent *P* trapping in the sediments, structural shifts in algal community towards fast sinking diatoms due to low *P* availability, nearshore organic matter that could be getting sequestered in offshore bottom sediments, and gradual increase of bioavailable phosphorus in the exogenous loading (Chapra et al., 2016). Following the *TPMB* evolution over the course of the past thirty years, there are several illustrations on how additional complexity can be accommodated by the present structure, while maintaining its parsimonious character (Yaksich et al 1985; Lesht et al. 1991; Gudimov et al., 2012; Shimoda and Arhonditsis, 2015; Katsev, 2017). On a final note, the predicted *TP* for each lake segment were causally linked with two trophic indicator variables, summer chlorophyll *a* and Secchi disk depth, based on the empirical relationships described in Dove and Chapra (2015); see their Figs 11 and 12. It is important to recognize though that (*i*) the two equations were derived by pooled data from all of the Great Lakes, (*ii*) the corresponding regression models explained between 55-60% of the observed variability, and (*iii*) the reported offshore summer chlorophyll *a* versus spring *TP* relationship distinctly underestimates the summer phytoplankton levels in the western Lake Erie basin.

1-Dimensional Central Basin Hypoxia Model: The *1D-CBH* model comprises a simple eutrophication model coupled with a *1-D* thermal model that provides temperature and associated vertical mixing profiles in the offshore waters of Lake Erie central basin (Rucinski et al., 2010; 2014). The *1-D* thermal model is based on the Princeton Ocean Model, which uses a Mellor-Yamada turbulence closure scheme to parameterize vertical mixing (Mellor and Yamada, 1982) and has been modified with an overland-overlake correction (Beletsky and Schwab, 2001). The eutrophication model explicitly simulates organic carbon, phytoplankton, zooplankton, and dissolved oxygen, while phosphorus is divided into two pools, i.e., available and unavailable (or particulate) *P*. The model is forced with carbon and phosphorus loading from both western and central basins, with the former loads routed to the central basin after accounting for a constant net apparent settling loss of 10 $m \cdot yr^{-1}$ (Lesht et al., 1991). Both hydrodynamic and eutrophication models operate with the same vertical resolution, i.e., 48 vertical layers (each 0.5 m thick). The model has been tested against 19 years (1987–2005) of observed loading rates and meteorological conditions, with the focal question being the relative contribution of thermal stratification conditions versus *P* loading magnitude and timing on the severity of hypoxia in the central Lake Erie Basin. Because the model does not include ice-formation processes, the hydrodynamic and eutrophication models were initialized each year, using the earliest cruise sample concentration, and thus each year is simulated separately, as opposed to a continuous 19-year simulation.

The *1D-CBH* model displayed excellent ability to capture the temporal evolution of the observed vertical temperature profiles (Rucinski et al., 2010), including the onset of stratification and thermocline development in summer. Maximum model error varied with depth, 1.9 ^{o}C and 3.4 ^{o}C for 1994 and 2005, while the overall *RE* was lower than 5%. Nonetheless, the *1D-CBH* tended to underestimate the sudden mixed layer depth increase in late August and early September caused by storms, as well as to the effects of horizontal advection and internal wave propagation that could not be reproduced by a *1D* model (Ruchinski et al., 2010). Dissolved oxygen was also reproduced fairly well, i.e., *RE*≈25%, *MEF*≈0.46), but the capacity to reproduce biological (phyto- or zooplankton) components, and other closely-related variables (*DRP*) is limited, i.e., *RE*>50%, *MEF*<0 (Ruchinski et al., 2014; Ruchinski et al., 2016). Overall, the model was suggestive that within the range of load seasonality observed from 1987 to 2005, the hypoxic response is far more dependent on stratification structure (i.e., early or prolonged stratification, deep thermocline) than on the timing of nutrient delivery into the system. The model consistently predicted that sediment oxygen demand (*SOD*) represents a substantial fraction of the overall oxygen demand. Nonetheless, it is important to understand that the *SOD*-versus-*TP* loading (rectangular hyperbola or simply a Michaelis-Menten-type) equation used represents a convenient means to run loading scenarios and directly evaluate the changes in water quality variables, but does not offer any mechanistic insights about the actual response of the sediment diagenesis processes nor does it have solid empirical foundation. After all, the precursor of that equation from Borsuk et al. (2001) was based on cross-sectional data from 34 estuarine and coastal systems worldwide, with the majority of those bearing little resemblance to Lake Erie (see Table 1 in Borsuk et al., 2001).

Ecological Model of Lake Erie: EcoLE is based on the two-dimensional hydrodynamic and water quality model *CE-QUAL-W2* (version 2.0), which was originally developed to simulate long and narrow waterbodies displaying longitudinal and vertical water quality gradients and lateral homogeneity (Cole and Wells, 2006; Zhang et al., 2008). *EcoLE* divided Lake Erie into 1 m vertical layers (maximum number 65 vertical layers) and 222 longitudinal segments from west to east. The model has six variables for hydrodynamic simulations: horizontal velocity, vertical velocity, free water surface elevation, pressure, density, and constituent concentration. The vertical eddy coefficient algorithm has also been modified to adjust the strength of seiches and improve the simulations of longitudinal currents in a large, wind-driven Lake Erie (Boegman et al., 2001). The ecological model includes 28 state variables, with flexibility to include more water quality variables (Cole and Wells, 2006). Phytoplankton was divided into non-diatom edible algae, non-diatom inedible algae, and diatoms, while zooplankton comprised cladocerans and four copepod variables (copepod eggs, nauplii, copepodites and copepods); the latter component was based on the stage-structured population model originally presented by Fennel and Neumann (2003). The ecological implications of dreissenids were considered by two processes: grazing on phytoplankton and nutrient recycling via excretion (Zhang et al., 2008). Both their clearance rate and excretion rates (*N* and *P*) were linked to the dreissenid population size per segment, which in turn was described as a function of depth-dependent density and sediment area (Jarvis et al., 2000), and basin-specific length-mass regressions (Zhang et al., 2008).

The hydrodynamics of the model were calibrated against the water levels, currents, and temperature data during the growing season (May-September) from 1994 (Boegman et al., 2001). The water quality module was subsequently calibrated using phytoplankton and nutrient concentration data from 1994 and 1997 (Boegman et al., 2008; Zhang et al., 2008), and subsequently validated against data from 1998 and 1999 (Zhang et al., 2008). The latter skill assessment exercise was mainly intended to evaluate the ability to capture the longitudinal patterns in Lake Erie, while the temporal variability was phased out by using seasonally averaged field measurements. The goodness-of-fit statistics showed that the model could not accurately reproduce the longitudinal variability in the system with *RE* varying from 30% to 120% and *MEF* values that were almost consistently negative. That is, *EcoLE* displayed inferior performance relative to the simple use of a lake-wide average value for each variable included in the calibration dataset (Zhang et al., 2008). Nonetheless, a recent model assessment against a snapshot of the vertical profiles for water temperature, *DO*, and chlorophyll *a* from the summer of 2008 provided more encouraging results, although some parameters (maximum sediment *DO* demand along with the associated half saturation, a coefficient in the cloud cover function) had to be readjusted in order to improve the representation of the vertical distribution of the water temperature and *DO* in the central basin (Scavia et al., 2016b). A key feature of the process characterization from *EcoLE* is that algal growth relies on *P* recycling within the upper water column, including crustacean excretion, organic matter decay, dreissenid mussel population excretion rates, while phosphorus release from the sediments occurred mainly in the central basin, when hypoxic conditions prevail. Similar to *1D-CBH*, the latter prediction is based on a simple engineering approach (user-specified sediment release rates) that does little in shedding light on the diagenesis process, nutrient retention time in the sediments, and their potential response to reduced sedimentation fluxes of autochthonous material (Scavia et al., 2016b).

Western Lake Erie Ecosystem Model: WLEEM was introduced by Verhamme et al. (2016) as a three-dimensional (*3D*), fine-scale modelling framework, developed to simulate water quality responses to changes in the meteorological conditions and discharges of nutrients and sediments from tributaries into the western basin Lake Erie. *WLEEM* comprises four process-based models: (*i*) *EFDC* (Environmental Fluid Dynamics Code), which is a *3D* Finite difference hydrodynamic thermodynamic model, originary developed by Hamrick (1992), providing water level and velocity as forcing functions for *SWAN*, and current direction and velocity as forcing functions for *SEDTRAN*. (*ii*) *SWAN* (Simulating WAves Nearshore), which simulates wave conditions based on wind, depth, frictions, and velocity. This submodel linked to the *EFCD* (input) and *SEDTRAN* (output) to provide wave forcing for circulation (i.e., wave height, period, and direction) and support the simulation of wind-driven resuspension. (*iii*) *SEDTRSN* (Sediment Transport Model), which is a *3D* sediment transport model developed based on *SEDZLJ* model by Jones and Lick (2001). This submodel offers flexible simulation options that provide sediment settling/deposition, resuspension/erosion, bed armoring, and *3D* transport of multiple cohesive and non-cohesive sediment classes. (*iv*) *A2EM* (Advanced Aquatic Ecosystem Model). The original modeling framework was developed for the Upper Mississippi River system by HydroQual, Inc. in the late 1990s (LimnoTech, 2009) and the same model was revised to add a finer representation of the plankton community to simulate their seasonal successional patterns. The water quality model *A2EM* was developed on a publicly available version of Row-Column *AESOP (RCA)* model code (Limnotech, 2010; 2013) that simulates carbon, nitrogen, phosphorus, and oxygen cycles. The model has the capacity to simulate five phytoplankton functional groups, three zooplankton functional groups, two dreissenid mussel size-based classes, and benthic algae (i.e., *Cladophora*) (Verhamme et al., 2016). The seasonal dynamics of only three major phytoplankton functional groups (i.e., diatoms or winter algal assemblage, green algae or summer algal assemblage, and cyanobacteria) were considered by Verhamme et al. (2016).

WLEEM was calibrated using monitoring data from March to November 2011 – 2013, which represent a wide range of environmental conditions in Lake Erie with both high and low flow, nutrient loads, wind patterns, and significant variability of nutrient concentration and algal biomass (Verhamme et al., 2016). Model calibration first focused on physical (hydrodynamics, temperature, suspended solids), followed by chemical (*TP* and *SRP*, nitrogen species) system components. The calibration of phytoplankton state variables revolved around the timing and peak magnitude of the simulated functional groups; namely, diatoms, greens, and cyanobacteria during spring, early summer, and middle to later summer, respectively. *WLEEM* successfully reproduced the temperature variability (RE<5% and MEF>0.830), and less so the ambient *TP* levels (43%<RE<53% and -0.949<MEF<0.067) in response to high-flow, spring-loading events, with concentrations being attenuated (diluted) with distance from the Maumee River mouth (Shimoda et al., 2018). *SRP* (53%<RE<75% and -0.054<MEF<0.448) and NO_2+NO_3 (32%<RE<41% and 0.351<MEF<0.726) concentrations were also reproduced reasonably well. Because of the limited data availability, the processes underlying the onset of spring (March–May) bloom were not adequately characterized. The model also tends to initiate *HAB* development earlier than the monitoring data suggest by approximately four weeks. Model performance against cyanobacteria (96%<RE<106% and -0.302 <MEF<0.097) was distinctly worse relative to chlorophyll *a* concentrations (58%<RE<71% and -0.315 <MEF<0.025).

Estuary and Lake Computer Model-Computational Aquatic Ecosystem Dynamics Model: *ELCOM-CAEDYM* is a coupled *3D* hydrodynamic (*ELCOM*) and biological (*CAEDYM*) model, and is considered one of the most commonly used models worldwide. *ELCOM* on its own has been used to examine the response of the thermal structure (Liu et al., 2014) and circulation patterns (Leon et al., 2012) in Lake Erie to changes in meteorological conditions (e.g., air temperature, wind speed). The same model has also been used to simulate *3D* transport of organic materials/ organisms, such as walleye larvae floating near the lake surface (Zhao et al., 2009). During its initial application, *ELCOM* has been calibrated to vertical temperature profile and data from 1994 for all basins, and subsequently against data from 2001-2003 in the eastern basin (Leon et al., 2005). Most recent simulations of the water temperature patterns displayed very satisfactory performance, i.e., *RE*<10% and *MEF*>0.780 (Bocaniov et al., 2014, Oveisy et al., 2014 Bocaniov et al., 2016, Bocaniov and Scavia, 2016).

The existing *ELCOM-CAEDYM* applications in Lake Erie opted for simpler ecological structure than its built-in capacity. Specifically, *CAEDYM* has a total of 112 state variables that consider water, sediment, chemical, biological processes, and can simulate up to 7 phytoplankton functional groups (*PFG*s), 5 zooplankton functional groups (*ZFG*s), 6 fish groups, 4 macroalgal groups, 3 invertebrate groups, 3 mussel classes, macrophytes, seagrass, jellyfish (Hipsey and Hamilton, 2008). The application of *ELCOM-CAEDYM* to Lake Erie is limited to five phytoplankton groups (i.e., early bloom diatoms, lake bloom diatoms, cyanophytes, cooler and deeper water flagellate and warmer and brighter water flagellates) (Leon et al., 2011; Bocaniov et al., 2014; 2016). State variables that represent zooplankton have never been activated with *ELCOM-CAEDYM* applications in Lake Erie, and thus the role of zooplankton has been accommodated solely by the increased grazing-induced losses of phytoplankton (Leon et al., 2011; Bocaniov et al., 2014; 2016). A novel feature of *ELCOM-CAEDYM* in Lake Erie is the explicit representation of dreissenids dynamics. The built-in dreissenid sub-model was developed based on a clam model for a temperate estuarine lagoon (Spillman et al., 2008). Dreissenid biomass density is modulated by grazing (on phytoplankton and detrital particle) and respiration, excretion, egestion, and mortality losses (Bocaniov et al., 2014). Nutrient release from the sediments and ice formation modules are also incorporated to *ELCOM-CAEDYM*. Both modules have been used to consider the role of sediments on the elevated *TP* in the northeastern nearshore zone (Leon et al., 2011), as well as to evaluate the effects of ice cover on winter productivity and subsequent hypoxia development (Oveisy et al., 2012).

ELCOM-CAEDYM was calibrated against *DO* observations (Leon et al., 2006). Another version with additional submodules (i.e., *PFG*s, dreissenids, and ice-cover) has been also recalibrated against temperature, chlorophyll *a*, light attenuation, and nutrient concentration data from 2002 (Leon et al., 2011; Bocaniov et al., 2014) and temperature, *DO*, ice-cover and thickness from the 2004-2005 winter period (October-April) data (Oveisy et al., 2014). The most recent studies (i.e., Bocaniov et al., 2016; Bocaniov and Scavia, 2016, Karatayev et al., 2017) improved the external forcing functions with finer resolution meteorological data and more comprehensive discharges from tributaries. The simulated thermal structure was validated against *USEPA* cruise data from 2008, satellite-derived, lake-wide surface observations (Bocaniov and Scavia, 2016), and *DO*, chlorophyll *a*, nutrient concentrations from two basin-wide cruises occurred in 2008 (Bocaniov et al., 2016). With the most recent studies, the reproduction of chlorophyll *a* concentrations was generally satisfactory with the skill assessment by Bocaniov et al. (2014), $18\% < RE < 42\%$ and $-0.467 < MEF < 0.868$, and less so with the one presented by Valipour et al. (2016), $64\% < RE < 88\%$ and $-3.081 < MEF < 0.250$.

Eastern Basin Cladophora Model: The *EBC* model is a simple mechanistic model aiming to predict *Cladophora* standing biomass and *P* stored in plant tissues (Auer et al., 2010). The original framework to simulate *Cladophora* growth/biomass developed by Canale and Auer (1982) has been modified by several authors (Higgins et al., 2005, 2006; Tomlinson et al., 2010; Auer et al., 2010), which generally focused on refinements of three terms: growth, loss by respiration and sloughing. The growth rate of *Cladophora* is expressed as a function of the maximum gross specific growth rate and limitation multipliers that account for the role of light, temperature, internal *P*, and maximum biomass density (carrying capacity). The respiration rate considers both a dark/basal rate that varies only with temperature as well as a light-enhanced rate determined by temperature and light intensity (Tomlinson et al., 2010). In its updated version, sloughing is modelled as a first-order loss process with the rate coefficient varying as a function of water temperature and the depth of colonization, which in turn reflects the effect of wind energy (i.e., momentum leading to detachment) and benthic shear stress effect. The physiological effects of temperature initiate with a minimum temperature until an optimum temperature level is reached at and above which the sloughing rate is at its maximum value (Tomlinson et al., 2010). In a recent study, Valipour et al (2016) incorporated *EBC* with *ELCOM-CAEDYM* in order to investigate the interplay among external phosphorus loading, nearshore phosphorus, and *Cladophora* growth. The introduction of *EBC* into a *3D* hydrodynamic environment intended to shed light on both nearshore and basin-scale physical factors, such as wind, surface runoff, degree of stratification, upwelling events (period of 5–10 days), seiches (~14 h), and near-inertial waves (~17 h) that shape offshore-nearshore exchanges. Upwelling events during the months of May-early July appear to be the predominant drivers of offshore-nearshore mass exchanges, whereby *SRP* is injected from the hypolimnion into the nearshore northeastern Lake Erie; especially since these events usually coincide with optimal light and water temperature conditions (Valipour et al., 2016).

Considering the limited empirical information, *EBC* displayed the skill to reproduce *Cladophora* biomass along the shallow areas of the eastern Lake Erie basin (<10m), but its performance declined at the deeper zone may be due to the inadequate parameterization of the sloughing mechanisms (Tomlinson et al., 2010). Our independent assessment of the goodness-of-fit against *Cladophora* biomass data showed that *MEF* varied considerably between -0.705-0.708 (Shimoda et al., 2018), while the median *RE* of all the "*observed-versus-predicted*" comparisons presented in the literature was larger than 60% (Fig. 5). In a similar manner, the corresponding values for the *P* tissue content varied from -2.795 to -0.016 (Shimoda et al., 2018). Except from our limited understanding of the drivers underlying the sloughing processes, some of the identified weakness include the lack of a comprehensive *SRP* database for nearshore habitat colonized by *Cladophora*, which inevitably leads to a reliance upon measurements from offshore sites or at water intakes located at depths often beyond the domain of colonization (Auer et al., 2010). Moreover, there is no empirical evidence of *Cladophora* biomass levels immediately prior to the dreissenid invasion, and thus there is no direct way to reconcile their role on *Cladophora* "resurgence" in Lake Erie (Higgins et al., 2005). In addition, the internal phosphorus is rarely measured in natural populations, and the values used typically span a wide range (Auer and Canale, 1982; Jackson and Hamdy, 1982; Higgins et al., 2005; Malkin et al., 2008; Thomlinson et al., 2010). Except from the lack of data, there also seems to be some inconsistency with respect to the sampling/analytical protocols followed during the determination of *Cladophora* biomass used to constrain the existing models.

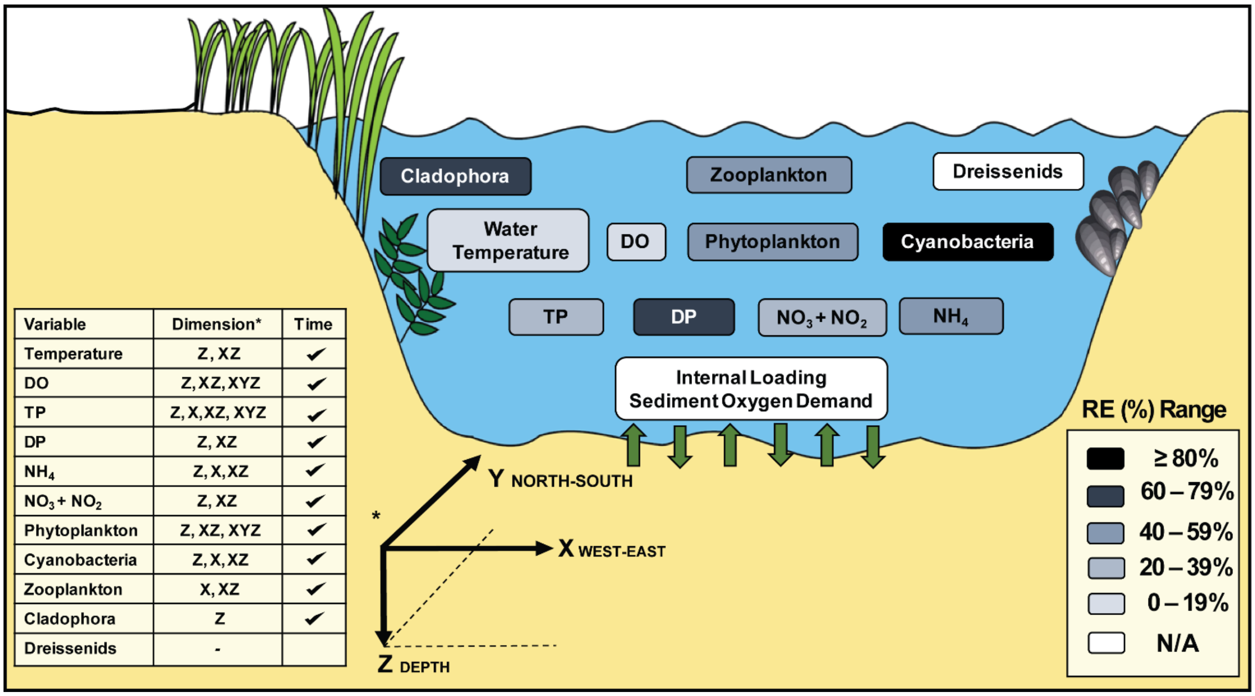


Figure 5: Model performance for different physical, chemical, and biological components of Lake Erie. The relative error for each state variable is based on the median value of all the corresponding graphs published in the peer-reviewed literature, in which field data were compared against simulated values. Counter to the practices followed in Scavia et al. (2016a,b), our goodness-of-fit assessment was based on point comparisons in time and space instead of aggregated (seasonal/annual, basin- or lake-wide) scales.

4 Lessons Learned from the Lake Erie Modelling Framework

4.1 *SWAT*-based Ensemble Strategy: Delineation of "Hot-Spots" in the Maumee River Watershed and Analysis of *BMP* Scenarios.

Following the development of a spatially distributed model, the identification of high-risk areas (or "hot-spots") with greater propensity for nutrient export and downstream delivery rates is an important exercise (Scavia et al., 2016c). In the Maumee River watershed, given that the calibration of all five *SWAT* applications was based on a single downstream station without any spatial information used to constrain the underlying processes, the (dis)agreement among the corresponding delineations could be primarily determined by two factors: (*i*) the discrepancies among the assumptions made or input data used during the spatial configuration (e.g., tile drainage, fertilizer/manure application rates, land use/land cover or *LULC* data) of the individual models; and (*ii*) the differences in the characterization of processes pertaining to the water and nutrient cycles. The accommodation of the latter source of uncertainty, i.e., the inability to unequivocally quantify the relative importance of major hydrological and nutrient transport/transformation mechanisms, is one of the main benefits of the *SWAT*-ensemble strategy in the Maumee River watershed. By contrast, the errors arising from the mischaracterization of boundary conditions and forcing functions (e.g., fertilizer application rates, agricultural tile drainage network) among the ensemble members represent "*nuisance*" factors that inflate the uncertainty of the predictions drawn without necessarily advancing our understanding of the watershed functioning or shedding light on the critical processes for achieving our environmental goals (Kim et al., 2018a).

Bearing these two major sources of uncertainty in mind, Scavia et al. (2016c) showed an agreement among the models in identifying higher *TP* loading rates from the northwestern and southern parts of the Maumee River watershed, whereas a tendency for higher dissolved reactive phosphorus (*DRP*) export rates was mainly projected on the predominantly agricultural central area (Fig. S1). Considering that a significant amount of *DRP* (45-60%) can be exported to rivers through agricultural tile drainage (Xue et al., 1998; Kovacic et al., 2000), this disconnect between *TP* and *DRP* loading is likely associated with the tile-drainage assumptions made during the spatial configuration of the five models. In particular, Kim et al. (2018a) showed that considerable variability exists among the individual estimates for surface runoff and tile flow (subsurface runoff), i.e., 191~275 mm and 50~139 mm, even though the total runoff (calculated as the sum of surface and tile flow, excluding groundwater) was fairly similar (301~334 mm) across the five

SWAT models (see Fig. 3 in Kim et al., 2018a). In the same context though, the differences in the spatial distribution of the tile-drainage among the models in conjunction with the fairly similar "hot-spot" mapping for a given phosphorus form (Fig. S1) may be an evidence of a retrofitting of the phosphorus-related parameters that counterbalances the differences in the simulated hydrology (see Table 2 in Kim et al., 2018a). For example, the *OSU* model, based on a parameter specification that postulated low tile flow and high mineralization rates, predicted similar *DRP* loading patterns in the south-central Maumee River watershed with the *TAMU* model, in which high tile flow was combined with low *P* mineralization rates (Kim et al., 2018). Given the latter evidence along with the fact that the calibration of all the models was based on a single downstream station, it can be inferred that the capacity of the current *SWAT* ensemble to pinpoint nutrient-export hot-spots is uncertain and may simply be the result of multiple errors that cancel each other out and ultimately lead to the same output. It is thus critical to mitigate the confounding effects of the nuisance factors of the current watershed modelling framework by improving and/or sharing identical input data (e.g., fertilizer application rates, *LULC* data, and agricultural tile drainage network). In doing so, we will establish a common denominator across all of the *SWAT* applications upon which the implications of other critical sources of uncertainty will be examined.

As previously mentioned, notwithstanding the fact that the uncertainty envelope derived from an ensemble strategy is potentially inflated by factors not directly related to the watershed response to alternative *BMP* scenarios, there are still compelling arguments in favor of their use for guiding environmental policy decisions (Scavia et al., 2016; 2017). In the Maumee River watershed, ten land-use management scenarios were designed after considering issues related to their practical implementation and policy feasibility, i.e., commonly applied (fertilizer reduction, tillage replacement) versus less frequent management practices (land-use conversions, wetland/buffer restoration); the ability of *SWAT* to examine certain agricultural activities; and extensive consultation with agricultural and conservation stakeholders (Scavia et al., 2016; 2017; see also details in Table S1). Overall, little evidence was provided regarding the likelihood to achieve the March-July phosphorus loading targets of 186 metric tonnes of *DRP* and 860 metric tonnes of *TP* (GLWQA, 2016), or 40% reduction from the 2008 loads, across the different *BMP* scenarios examined (Figs. S2a,b). Interestingly, the attainability of the *TP* loading threshold seems to be more likely relative to the one for *DRP* loading. It is also worth noting that the forecasts associated with commonly applied *BMP* scenarios (*S1*-*S4*) were somewhat more conservative, in comparison with scenarios that are less frequently applied (*S5* and *S9*) (Figs S2a,b; see also discussion in Kim et al., 2018a). Moreover, the degree of divergence of the individual model forecasts for the various *BMP* scenarios examined (or the forecasting spread) offered insights that can meaningfully influence the environmental policy analysis process. In particular, the forecasting spread increases significantly with the degree of deviation of *BMP* scenarios from the present conditions (Figs S2c,d). The existing *SWAT* applications suggest that for every 50 metric tonnes of reduction achieved the standardized forecasting spread, or the deviation of the five models divided by their corresponding averaged prediction for a given scenario, increases by 1.5% and 13% for *TP* and *DRP*, respectively (see linear regression equations in Figs S2c,d). To put it another way, the same scatterplots suggest that the standardized forecasting spread with the existing loading targets of 860 tonnes for *TP* and 186 tonnes for *DRP* is 26% and 33%, respectively.

The likelihood of a moderate reduction of *DRP* loading with nearly all the different management practices examined is consistent with recent empirical and modelling evidence from the Lake Erie watershed (Daloglu et al., 2016; King et al., 2017; Baker et al., 2017). In particular, the increasing *DRP* loading trend after the mid-1990s has been attributed to the increased frequency of storm events, suboptimal fertilizer application rates and timing, and management practices that appear to increase phosphorus accumulation at the soil surface (Daloglu et al., 2016). Counter to the (nearly) monotonic decline of the amount of nitrate in soils, which exhibited significant vertical mobility and was distinctly flushed out of the watershed after rainfall events, *DRP* appears to display a remarkable persistence across a wide range of spatial scales (King et al., 2017). The latter pattern highlights the role of legacy *P* as an important regulatory factor of the *DRP* concentration in soils. In particular, historical *P* fertilizer application rates seem to have led to soil reserves well-above critical levels in the area (Powers et al., 2016), which allowed to maintain high crop yields even after *P* fertilizer implementation declined (Dodd et al., 2013; Liu et al., 2015; McDowell et al., 2016). *P* build-up is hypothesized to have triggered a rapid, microbially mediated transformation of labile to bioavailable pools that are more susceptible to losses (King et al., 2017). If the latter mechanism holds true, then surface fertilizer applications could exacerbate soil *P* stratification. While this possibility casts doubt on the management practices designed to reduce top-soil *P* levels, the periodic soil inversion tillage to thoroughly mix the soil in the plow layer is recommended as the best strategy to avoid stratification increments (Baker et al., 2017). Moreover, alongside the buildup of legacy labile *P* fractions at the soil surface, recent research suggests the establishment of macropore flow pathways and increased tile drainage facilitate the hydrological connectivity and ultimately transport of soluble *P* among soil surface, subsurface drainage, and stream network (Sharpley et al., 2011; Jarvie et al., 2017). Given this emerging evidence, one challenging aspect for the evaluation of scenarios with the *SWAT*-ensemble is the proper consideration of legacy *P* (e.g., initialization that accommodates the spatial soil *P* variability, sufficient model spin-up period, parameter specification that reproduces the gradual *P* accumulation in the soils) and its ability to reproduce the critical hydrological and transformation mechanisms modulating the *DRP* loading in the Lake Erie basin.

4.2 Lake Erie Multi-Model Ensemble Strategy: Distinguishing Between Predictable Patterns and Sources of Uncertainty with the Load-Response Curves.

In this section, we critically evaluate the credibility of the Scavia et al. (2016a,b) forecasts on the achievability of the four *ERI*s, after forcing the Lake Erie multi-model ensemble with a series of nutrient loading reduction scenarios. To impartially conduct this analysis, we believe that these predictive statements cannot be viewed independently from the technical features of each of the modelling tools used, the characterization of the fundamental ecological processes modelled, and the ecological insights gained. In the same context, it is important to note that there is no consistent information from the individual members of the model ensemble regarding the quantification of all the fluxes pertaining to modelled biogeochemical cycles (carbon, nitrogen, phosphorus, oxygen), and therefore it is difficult to evaluate the relative significance of critical ecological pathways (e.g., internal loading, importance of bacterial-mediated nutrient regeneration or nutrient excreta/egesta from zooplankton/dreissenids, seasonality of sedimentation fluxes) that can conceivably modulate the projected response of Lake Erie to exogenous nutrient loading reduction. In the next iteration of the modelling framework, we thus highlight the need to report for each ensemble member the resulting quantitative description of the simulated biogeochemical cycles (see Fig. 7 in Kim et al., 2013 or Fig. 4 in Gudimov et al., 2015), which in turn will not only provide an essential piece of information to put the derived ecological forecasts into perspective, but will also allow to understand how distinctly different are the characterizations of the ecosystem functioning and consequently the degree of "diversity" of the modelling tools at hand. We should always bear in mind that the actual benefits of using multiple models do not stem from the complexity of the mathematics per se, but rather the ability to test alternative ecosystem conceptualizations or competing hypotheses regarding the role of certain facets of the underlying biogeochemistry.

4.2.1. Basin-specific overall phytoplankton biomass represented by summer average chlorophyll-a concentrations: Existing evidence from the international peer-reviewed literature suggests that the current generation of process-based models has satisfactory ability to reproduce the observed patterns of total phytoplankton biomass in a wide variety of aquatic systems and trophic conditions (Arhonditsis and Brett, 2004; Shimoda and Arhonditsis, 2016). Chlorophyll *a* as a bulk estimate of the total phytoplankton biomass across all the functional groups is a commonly used criterion to assess the trophic status of lakes. Lake Erie historically exhibits the highest lake-wide chlorophyll *a* concentrations with a distinct longitudinal west-east gradient (Cai and Reavie, 2018). Oligotrophic conditions prevail in the offshore waters of the central and eastern basins, rarely exceeding an average summer chlorophyll *a* concentration of 2.5 μg Chl *a* L^{-1} over the last three decades (Dove and Chapra, 2015). On the contrary, seasonal average chlorophyll *a* levels in western basin often exceed the mesotrophic levels with the highest concentrations typically experienced during early fall (Zolfaghari and Duguay, 2016). Hence, the focus of the analysis of nutrient loading reduction scenarios and the development of load-response curves was on the western basin (Scavia et al., 2016a,b). The models used for the latter exercise were very diverse, including an empirical model coupled with *TPBM* and three complex process-based models (*EcoLE*, *WLEEM*, and *ELCOM-CAEDYM*) (Scavia et al., 2016a,b). The load-response curves produced from the four models displayed distinct differences in the predicted trajectories of the system. For example, Scavia et al. (2016b) noted that the annual *TP* loads into the western basin needed to bring about a 50% decrease in the maximum chlorophyll *a* concentration ranged between 1,130 MT and 3,010 MT (Fig. S3a). The substantial uncertainty band raises the question to what extent it stems from the diversity of the models used and their conceptual/structural differences in terms of the suite of ecological drivers considered (i.e., the "desirable" variability that a model ensemble aims to accommodate) or from mischaracterizations of the system ecology and their individual biases that unnecessarily inflate the uncertainty of the model ensemble.

One of the members of the Lake Erie ensemble was based on a chlorophyll *a* vs *TP* exponential regression model derived by data pooled from all of the Great Lakes (Dove and Chapra, 2015). As previously mentioned, this empirical equation apparently underestimates the summer phytoplankton levels in the western Lake Erie basin, which likely stems from several lower-end (Lakes Huron and Michigan), and mid-range (central and eastern Lake Erie) observations that shape its intercept and slope, respectively (see Fig. 11 in Dove and Chapra, 2015). Building regression models from cross-sectional data is not an unusual practice in limnology, and this bias can be easily rectified by formulating a hierarchical structure that relaxes the assumption of globally common parameters and allows estimating system-specific regression coefficients

(Cheng et al., 2010). It is also important to consider other predictors that could conceivably influence the levels of total phytoplankton biomass, given that the existing regression model explained less than 60% of the observed variability. We emphasize that it is essential to include and suitably update (at least) one data-driven model per *ERI* considered, as their empirical foundation offers a distinct alternative to the "growth-minus-loss" mass balance strategy characterizing the phytoplankton governing equation of process-based models.

The phytoplankton growth terms of *EcoLE* and *ELCOM-CAEDYM* bear conceptual resemblance in that they both explicitly consider the limitations posed from temperature, whereas all the other potentially limiting factors (e.g., light, nutrients) are included within the minimum formula of the Liebig's Law and thus only one of those actively limits algal growth in a given time step. Regarding the plausibility of this strategy, Kim et al. (2014) argued that likely overstates the role of light availability as the predominant factor of the bottom-up forcing in the system, given that the low values assigned to half-saturation constants for phosphate uptake could be reducing the severity of phosphorus limitation in the summer. Counter to this assertion though, Zhang et al. (2016) showed that *EcoLE* simulations are suggestive of years (1997) when phosphorus is the sole limiting factor, even under the present (baseline) conditions, and other years (1998) when light and phosphorus interchangeably determine the degree of algal growth limitation during the summer period (see Fig. 11 in Zhang et al., 2016). According to forecasts drawn from *EcoLE*, it is also important to note that the initial benefits (i.e., decrease of total phytoplankton biomass) from the implemented nutrient loading reduction will emerge from the residents of the summer algal assemblage that possess inferior kinetics for *P* uptake (i.e., their non-diatom inedible algal group), whereas other species/superior *P* competitors (diatom-like or non-diatom edible algae) will display minimal decline (if not increase) because the severity of *P* limitation will be offset by the resulting improvement in the light environment (Zhang et al., 2016). The same pattern of a moderate response at the initial stages of loading reduction becomes more evident with the *WLEEM* load-response curve (see pages 14-15 in Scavia et al., 2016b), as the likelihood of a counterbalancing effect between *P* and light limitation carries more weight with the corresponding simulations, given that light availability is treated as an independent limiting factor (Verhamme et al., 2016). By contrast, the *ELCOM-CAEDYM* load-response curve is less steep than the rest of the models considered, even though the mathematics of the phytoplankton growth limitation term are fairly similar to *EcoLE*. One plausible explanation could be that *P* limitation with this model is based on a two-pronged process that first considers the nutrient uptake rate in relation to the ambient nutrients and subsequently the growth rate as a function of their internal nutrient storage. The latter strategy postulates that the intracellular *P* pool buffers the phytoplankton response to nutrient variability. Although we cannot unequivocally pinpoint which of the three representations of the same process (phytoplankton growth) is the *"correct"* one for Lake Erie, we believe that this is an excellent example of how multiple competing models can capture an influential source of uncertainty.

The characterization of the rest of the algal processes (respiration, excretion, senescence mortality, sedimentation rate) was fairly similar, and therefore their influence on the uncertainty of the load-response curves must have been negligible. One distinct structural difference among the mechanistic models was that only *EcoLE* explicitly considered the importance of zooplankton dynamics in Lake Erie. The limited consideration of the role of zooplankton is somewhat surprising, given the general attention to the mechanisms underlying the strength of trophic coupling in planktonic food webs and the local work on its diurnal and seasonal adaptive behavior to overcome the hypolimnetic hypoxia (Vanderploeg et al., 2009). In its place, the existing modelling work focused on the role of dreissenids as the primary factor modulating phytoplankton abundance and composition, which likely holds true in the nearshore zone (Boegman et al., 2008; Zhang et al., 2008). However, improvement of zooplankton representation in the next iteration of Lake Erie modelling is important not only because it will allow to more effectively account for the likelihood of top-down control in offshore waters, but will also shed light on the broader implications of empirical evidence that certain residents (microzooplankton) of the zooplankton assemblage are more resilient than others (mesozooplankton) to toxic cyanobacterial blooms (Davis et al., 2012). In particular, an emerging hypothesis is that microzooplankton communities display a greater potential to serve as a top-down regulatory factor of toxic cyanobacterial blooms, whereas the inhibition of the mesozooplankton grazing rates by unpalatable algae could decrease the efficiency of carbon transfer and thus the upper trophic level productivity (Davis et al., 2012).

Expect from the planktonic processes, another significant driver of the trajectories predicted from the different models may be the postulated reliance of phytoplankton growth upon internal nutrient sources, which in turn can modulate the projected impact of external nutrient loading variability on standing algal biomass (Gudimov et al., 2011). Namely, a comparison of the intercepts of the load-response curves from the three mechanistic models, either as a scaled (%) phytoplankton response to its maximum (Fig. S3a) or as an actual chlorophyll *a* concentration (see graphs in pages 14 and 15 of Scavia 2016b), is indicative of distinct differences in the predicted levels of standing phytoplankton biomass in the western Lake Erie, even when external loading is practically eliminated (*ELCOM-CAEDYM* versus *EcoLE* or *WLEEM*). One possible reason could be the mineralization of organic compounds by heterotrophic zooplankton and microbes, which represent a major nutrient source to fuel phytoplankton production during the period of summer stratification and low ambient nutrient availability in the pelagic zone (Vanni, 2002; Teubner et al., 2003; Kamarainen et al., 2009; Ramin et al., 2012b). Earlier work from Goldman (1984) has described the intense microbially mediated regeneration as a rapidly turning "spinning wheel" by which nutrients are returned into the water column in short time scales (<1 day) with minimal losses. There is abundant evidence from a wide range of morphologically and geographically diverse lakes that the excretion of inorganic phosphorus by zooplankton can potentially account for a significant fraction of the phytoplankton demands (Gulati et al., 1995; Arhonditsis et al., 2004; Conroy et al., 2005; Kowalezewska-Madura et al., 2007). *EcoLE* was the only member of the model ensemble that explicitly considered the importance of this mechanism, suggesting that crustacean excretion may provide between 20-25% of the algal uptake (Boegman et al., 2008; Zhang et al., 2016). In the same context, dreissenid excreted/egested material could represent another major internal source of *P*, and existing empirical and modelling evidence suggests that its contribution could be responsible for a significant fraction (15-30%) of the phytoplankton uptake in the western Lake Erie, especially when external loads are low (Mellina et al., 1995; Arnott and Vanni, 1996; Conroy et al., 2005; Zhang et al., 2016). Considering that the importance of this source will likely increase with the proliferation of dreissenid mussels in the western basin (Karatayev et al., 2014), increasing temperature and/or prolonged stratification (Ramin et al., 2012b; Johengen et al., 2013), it is critical to quantify the actual role of the corresponding internal nutrient fluxes to the overall *P* budget and revisit the existing models accordingly. Both *ELCOM-CAEDYM* and *WLEEM* have submodels that are specifically designed to accommodate the role of dreissenids

Another important factor that could determine the response of Lake Erie to nutrient mitigation strategies is the likelihood of internal loading from the sediments. A recent study by Matisoff et al. (2016) attempted to shed light on this issue by obtaining estimates of the *P* diffusive fluxes from bottom sediments throughout the western basin of Lake Erie. A first striking finding was that the annual aerobic *P* flux for the entire a western basin is 378 MT P/year with a 95% confidence interval of 359 and 665 MT P/year, which in turn could correspond to anything between 11-60% of the recommended target loads of 1,130-3,010 MT P/year in order to achieve a 50%

reduction in maximum chlorophyll *a* concentration reported by each model used (see Table 12 in Scavia et al., 2016b). In a similar manner, depending on the temperature correction factor used to calculate the aerobic *P* fluxes and under the assumption of complete mixing with an average water residence time of 50.7 days, Matisoff et al. (2016) estimated that the sediment contributes between 3.0 and 6.3 μg L^{-1} of dissolved *P* to the water column, which represents 20–42% of the *IJC* Target Concentration of 15 μg P L^{-1} for the western basin. It is also interesting to note that a recent study from the central basin reported *P* release from the sediments that could be up to 20% of the total external input of *P* to Lake Erie (Paytan et al., 2017). Moreover, Matisoff et al. (2016) found that the anaerobic conditions are (on average) 4 to 13 times larger than those under aerobic conditions. The greater flux under anaerobic conditions were likely related to microbial reduction of *Fe* under anaerobic conditions, desorption of *P* into porewater, and subsequent diffusion into the overlying waters (Dittrich et al., 2013). While aerobic conditions at the sediment-water interface probably dominate redox chemistry in the western basin because of the shallow and mixed environment, the anaerobic fluxes may still be relevant since the shallow waters of the western basin can occasionally stratify for short time periods of up to 4-5 days and the bottom waters can become anaerobic.

It is thus critical to understand the intensive microbiological, geochemical, and physical processes occurring within the top few centimetres of the sediment and determine the fraction of organic matter and nutrients released into the overlying water (Table S2). Diagenetic modelling is an indispensable tool to investigate the interplay among the sediment processes, to verify concepts, and to potentially predict system behaviors (Gudimov et al., 2016; Doan et al., 2018). This kind of diagenetic modelling as well as the data that necessitate to ground-truth those models (e.g., porewater analysis, phosphorus fractionation, organic matter profiles) are still missing in Lake Erie; especially from the central basin. Field, experimental, and modelling work should be designed to shed light on the mechanisms of phosphorus mobilization in the sediments and to identify process controls under a variety of conditions. Interestingly, although it cannot offer the mechanistic insights (e.g., primary and secondary redox reactions, mineral precipitation dissolution reactions, acid dissociation reactions, and *P* binding form reactions) that the new generation of sediment diagenetic models can, *WLEEM* has a potentially useful sediment mass-balance submodel that distinguishes between aerobic and anaerobic layers and uses linear partitioning to assign the total amount of a nutrient between dissolved and particulate fractions (Verhamme et al., 2016). A similar engineering-type approach is also incorporated to *ELCOM-CAEDYM* (Hipsey, 2008). Both submodels can provide useful tools to dynamically track the response of the sediments under different nutrient loading regimes as long as several influential parameters (e.g., burial rates, particle mixing velocity among layers, resuspension rates, and aqueous mass transfer coefficients) are realistically constrained by empirical information from Lake Erie.

4.2.2. Cyanobacteria blooms in the western Lake Erie represented by the maximum 30-day average cyanobacteria biomass: For the purpose of establishing nutrient loading targets for this *ERI*, a 30-day average cyanobacteria biomass metric was selected as a proxy of the severity of harmful algal blooms in the western basin. In particular, a threshold of 9,600 MT dry weight biomass has been selected to distinguish between "severe" and "mild" blooms (Stumpf et al., 2012), based on existing records of satellite-estimated peak 30-day blooms since the early 2000s. Three models have been used to draw ecological forecasts and create load-response curves, i.e., *UM/GLERL* Western Basin *HAB* model, *NOAA* Western Basin *HAB* model, and *WLEEM* (Figs. S3b-d), but it should be noted that both *ELCOM-CAEDYM* and *EcoLE* consider multiple phytoplankton functional groups and could potentially be included in a similar exercise in the future. The synthesis of the predictions from two empirical equations and one mechanistic model has been somewhat problematic due to the different methods used to determine peak 30-day average cyanobacteria biomass. The former models are based on the satellite-derived estimates of maximum 30-day average bloom size calculated from consecutive 10-day composite images, which are in turn obtained by summing the highest biomass values observed at each pixel over each 10-day period (Stumpf et al., 2012), whereas *WLEEM* derives a maximum 30-day moving average from basin-wide daily simulations of cyanobacteria biomass (Verhamme et al., 2016). To reconcile this mismatch, a *WLEEM* prediction of 7,830 MT has been used as the equivalent to the satellite-derived bloom size of 9,600 MT.

Notwithstanding their simplicity and small sample size used for their development, the consideration of two empirical models for the assessment of this particular *ERI* is critical, as evidence from the modelling literature provides little support to the ability of the current generation of mechanistic models to forecast structural shifts in the composition of phytoplankton assemblages and patterns of cyanobacteria dominance (Anderson, 2005; Shimoda and Arhonditsis, 2016). In particular, a recent meta-analysis of 124 aquatic biogeochemical models found moderate fit statistics against empirical cyanobacteria biomass estimates, i.e., n=70, median r^2 = 0.36, RE = 65%, MEF = 0.06 (Shimoda and Arhonditsis, 2016). Of equal importance is the fact that there is considerable uncertainty with respect to the characterization of the ecophysiological traits of a "cyanobacteria-like" group, although there is a tendency to be defined as *K* strategists, displaying slower growth and metabolic rates that offer superior competitive skills in environments reaching their carrying capacity, with inferior *P* and superior *N* kinetics, adaptive capacity to tolerate turbid waters with low light availability, aptitude to sink slowly or even to regulate their vertical position within the water column in order to exploit favorable micro-environments, and limited preference (or even selective rejection) from zooplankton and dreissenids (Shimoda and Arhonditsis, 2016). The mechanistic modelling work that is in place in Lake Erie does not deviate from this general delineation of cyanobacteria, but there is considerable variability with the actual values assigned to the individual ecophysiological parameters (Zhang et al., 2016; Verhamme et al., 2016; Bocaniov et al., 2016). The only notable discrepancy from these practices was *WLEEM*'s characterization of cyanobacteria as the fastest growing group of the simulated phytoplankton assemblage, which in turn may explain the steep slope (8.37 MT of cyanobacteria biomass per MT of spring *TP* load from the Maumee River) of the corresponding load-response curve (see Fig. 12 in Verhamme et al., 2016). Overall, the load-response curves of the three members of the model ensemble coalesced with respect to their predictions, suggesting that bloom sizes below the selected threshold can be achieved with cumulative Maumee March–July loads of 890–1150 MT or annual loads of 1679–2170 MT (see Table 2 in Scavia et al., 2016a). The questions arising though is do the models converge for the right reasons or better yet what is the degree of our confidence on these predictions?

If the skill assessment is one of the criteria to infer about the credibility of the existing *HAB* process-based modelling work, then our view differs from the practices followed in Lake Erie with respect to what constitutes an acceptable model fit or even an appropriate temporal scale to evaluate model performance. Namely, Scavia at al. (2016b) reported a -14% percent bias and 38% mean absolute relative error for *WLEEM*, when seasonally averaged cyanobacteria biovolume data were compared against the corresponding model outputs across four stations. Although this spatially and temporally aggregated assessment was intended to evaluate performance against the proposed metric for this *ERI*, abrupt and non-linear compositional shifts in phytoplankton assemblages are the very essence of *HAB*s, and as such their study should be focused on finer time scales in order to understand the underlying mechanisms as well as their potential timing or actual magnitude. In particular, a careful inspection of the graphs included in Verhamme et al. (2016) lends little support to the ability of the model to predict cyanobacteria blooms (see their Fig. 8) and the reported goodness-of-fit statistics were clearly inflated from sites that did not exhibit significant cyanobacteria biovolume increase (e.g., station *GR1*) during the study period. In fact, our independent assessment of the model fit with daily resolution was suggestive of a $RE \approx 100\%$ and $MEF < 0$. Taken together with the substantial residual variability of the two empirical models, it can be inferred that the load-response curves for this *ERI* are surrounded by considerable predictive error and the relatively narrow uncertainty bounds are likely misleading.

Viewed from a heuristic perspective though, the existing cyanobacteria modelling work has collectively advanced our understanding of the factors triggering *HAB* formation. Considering that cyanobacteria dominance is largely the outcome of the resource competition among multiple phytoplankton species, one important lesson learned from both mechanistic and data-driven models was that both the dissolved reactive and particulate fractions of *TP* load must be taken into account when setting *HAB*-related load targets (Bertani et al., 2016; Verhamme et al., 2016). Existing empirical estimates show significant variability of the bioavailable fraction of particulate phosphorus (20-45%) in the Maumee River, and several mechanisms (e.g., microbial mineralization, anoxic release from the sediments) could potentially determine the bioavailability of the inflowing material from the time of entry in early spring until the mid-summer initiation of *Microcystis* blooms (Bridgeman et al., 2006; Loewen et al., 2007; Stow et al., 2015; Ho and Michalak, 2015). As previously mentioned, another interesting finding from the modelling work in Lake Erie is that its susceptibility to *HAB* occurrence could be increasing, and this trend could be attributed to changing meteorological conditions, such as warmer temperatures and calmer summer conditions (Michalak et al., 2013). Nonetheless, the signature of climatic forcing is not always evident on the timing and magnitude of *HAB* occurrence, and therefore a suite of alternative mechanisms have been proposed to explain the likelihood of an increase in the frequency of cyanobacteria dominance in the summer assemblage of Lake Erie, including the presence of an increasing reservoir of *Microcystis* seed colonies (Rinta-Kanto et al., 2009) and the selective filtering of dreissenids on competing phytoplankton species (Vanderploeg et al., 2001). Another factor that has received little attention is the importance of the inter-specific competition for various nitrogen forms; in particular, urea and ammonium are considered energetically favorable forms for protein synthesis and therefore predominant stimulants of *Microcystis* blooms (Finlay et al., 2010; Chaffin and Bridgeman, 2014). There is emerging evidence from other locations around the Great Lakes of a positive relationship between nitrogen concentration and toxin-producing *Microcystis* strains (Murphy et al., 2003) or microcystin production (Orr and Jones, 1998; Shimoda et al., 2016a; Kelly et al., 2018). Research also pinpoints to iron availability as another potential factor in triggering *HAB* events (Molot et al., 2014) or even herbicides/pesticides and other persistent organic pollutants (Harris and Smith, 2016), but little work has been done to evaluate this hypothesis in Lake Erie.

Another major unknown is the potential role of intense precipitation events in initiating c*HAB* events in Lake Erie. Our evolving understanding of their role in lakes identifies two classes of effects of weather events on abiotic conditions: short-lived effects of storms on lake thermal structure, and more prolonged effects of precipitation events on nutrient levels and water clarity (Droscher et al., 2009). The abrupt abiotic changes associated with extreme events can subsequently trigger changes in biotic ecosystem components (e.g., primary productivity, composition of plankton assemblages), but the magnitude of these shifts is predominantly modulated by the lake trophic status as well as fundamental system attributes (e.g., fetch, depth, water levels, wind regimes) (Shimoda et al., 2011). Consistent with these working hypotheses, Zhang et al. (2016) argued that even though the reduction in external phosphorus would result in distinct decline in algal biomass and *Microcystis* blooms in the western basin, several phosphorus input pulses stemming from storm events could induce favorable conditions for net *Microcystis* growth and ultimately dominance. Specifically, storm runoff events can bring into the system a substantial amount of nutrients but also elevate the turbidity, either through riverine input or sediment resuspension, which creates an environment that renders competitive advantage to *Microcystis* over the rest residents of the summer phytoplankton assemblage (Belov and Giles, 1997; Harke et al., 2016). Along the same line of evidence, Stumpf et al. (2016) asserted that the inflows from Maumee River not only profoundly influence the abiotic environment, but also combined with the prevailing circulation patterns can determine the exact area in the western basin where *HAB*s may initiate (Schwab et al., 2009). Thus, one of the important augmentations of the existing modelling framework is the establishment of direct linkages between watershed processes (as characterized by the *SWAT*-ensemble) and the receiving waterbody, in order to advance our understanding on the capacity of the perturbations induced by extreme runoff events to shape the complex interplay among physical, chemical, and biological components in western Lake Erie.

4.2.3. Central basin hypoxia represented by number of hypoxic days; average extent of hypoxic area during summer; and average hypolimnion DO concentration during August and September: Being different manifestations of the hypoxia in the central basin of Lake Erie, the three metrics were intended to offer a comprehensive assessment of the magnitude of the problem in order to capture its broader ramifications on ecosystem integrity. Specifically, the threshold for the average August-September hypolimnetic minimum *DO* concentration was originally set at 2.0 mg L^{-1}, but a level of 4.0 mg L^{-1} was ultimately selected, as it offers a more sensitive proxy to discern the onset of hypoxic areas (Zhou et al., 2013). The targeted levels for the average extent of hypoxic area during summer and number of hypoxic days were subsequently set to 2000 km^2 and 30 days, respectively. Counter to the practices followed with the previous two *ERI*s, the annual total phosphorus loads from tributaries to both western basin and central basin were identified as a better predictor for the hypoxia problem. Three models were used to generate the load-response curves with differences in the spatial resolution of the predictive statements drawn, i.e., *1D-CBH* model provided horizontally-averaged *DO* values, whereas *ELCOM-CAEDYM* and *EcoLE* horizontally-resolved *DO* concentrations in the bottom layer (0.5–1.0 m for *ELCOM-CAEDYM*; 1.0 and 1–3 m for *EcoLE*). In addition, *1D-CBH* model was forced with two distinct specifications of the load transport from the western to the central basin using: (i) the default net apparent attenuation loss rate due to settling; and (ii) the daily outputs from *WLEEM* to account for the mass fluxes crossing the western-central basin boundary. Regarding the achievability of the 4.0 mg *DO* L^{-1} threshold, the load-response curves provided a fairly wide uncertainty range, 2,600–5,100 MT, within which this target can be realized (Fig. S3e). Similarly, a load reduction anywhere between 3,415-5,955 MT was projected to reduce the average hypoxic extent to 2000 km^2 and the number of hypoxic days between 9 to 42 days (Scavia et al., 2016a,b).

The response curves for August–September average hypolimnetic *DO* concentration displayed similar increasing trends with decreasing loads, while some discrepancies arose at lower nutrient loading conditions; especially when the threshold *DO* level of 4 mg L^{-1} was reached (Fig. S3e). Although the latter uncertainty was partly attributed to the differences of the *SOD* formulations (Scavia et al., 2016a), the reality is that there were no significant conceptual differences among the three models used. The discrepancies were mainly introduced by the specific assumptions made to parameterize simple approximations of the potential response of the sediments to the reduced nutrient loading. In particular, as previously mentioned, *1D-CBH* used an empirical equation to connect *TP* loading directly to *SOD* based on a cross-sectional dataset of 34 estuarine and coastal systems, which was then used to predict the response of Lake Erie to loading reductions under the assumption that the large-scale (cross-sectional) patterns described in the model are also representative of the dynamics of individual systems (Cheng et al., 2010). In doing so, we essentially assume that all the systems in the dataset have identical behavior and therefore the empirical relationship is the same among and within systems. Consequently, after fitted to the cross-sectional dataset, Rucinski et al.'s (2014) equation provides a *SOD* value equal to 0.50 g O_2 m^{-2} d^{-1} when the annual *TP* loads are 4000 MT year^{-1}, which in turn is predicted to bring ambient *DO* close to 4 mg L^{-1}. On the other hand, *EcoLE* expressed *SOD* as a function of temperature and oxygen concentration (the latter relationship was mathematically described by a rectangular hyperbola), with an oxygen half-saturation constant for *SOD* set equal to 1.4 mg O_2 L^{-1} and a maximum *SOD* rate at 20°C (0.22 g O_2 m^{-2} d^{-1}) that was consistent with the range reported by Smith and Matisoff (2008; see their Table 1). Because of the values assigned to the two parameters, the oxygen losses to the sediments were predicted to be much lower (≈0.16 g O_2 m^{-2} d^{-1}) than *1D-CBH* when *DO* exceeds the level of 4 mg O_2 L^{-1}, which led to a disproportional *DO* increase when the external loads are low, i.e., <2000 MT year^{-1} (see Fig. S3e). Simply put, *1D-CBH* and *EcoLE* estimate distinctly different *SOD* rates (0.50 vs 0.16 g O_2 m^{-2} d^{-1}) under the same level of nutrient loading. Similar to the chlorophyll *a* predictions, *ELCOM-CAEDYM* resulted in a load-response curve that is less steep than the rest of the members of the model ensemble, which likely stems from the fact an even higher maximum *SOD* rate at 20°C (1.2 g O_2 m^{-2} d^{-1}) was assumed for this exercise (Bocaniov et al, 2016). While the same model postulates that the maximum *SOD* rate decreases with decreasing *TP* loads, the estimated *SOD* fluxes are much higher than any other of the models used (see Table S4 in Bocaniov et al., 2016). Based on the *SOD* rates reported by Smith and Matisoff (2008), the values used by *EcoLE* appear to be closer to the recent trends in Lake Erie, but the likelihood of higher oxygen demand rates cannot be ruled out given the considerable spatial and temporal variability characterizing the system (Matisoff and Neeson, 2005; Schloesser et al., 2005; Paytan et al., 2017).

As previously mentioned, to effectively augment the diversity of models included in the Lake Erie ensemble, there is a need for a sediment diagenesis model that quantifies *SOD* as the sum of the organic matter mineralization processes and the oxidation of reduced substances within different sediment layers (Gudimov et al., 2016). Only the characterization of the redox-controlled processes and the determination of the vertical profiles of biodegradable organic matter can allow a reliable estimation of the mineralization half-life period and subsequently the response rates of the sediments following different management actions. It is important to recognize that none of the existing members of the model ensemble has the mechanistic foundation to account for the likelihood of a lagged sediment response and unexpected feedback loops (Smith and Matisoff, 2008). Along the same line of evidence, the meteorological forcing and its effects on the thermal structure (e.g., timing and duration of stratification, thickness of hypolimnion) along with the water level fluctuations in Lake Erie are two important confounding factors that could further broaden the uncertainty band around the load-response curves predicted by the multi-model ensemble (Watson et al., 2016). Both theoretical and empirical evidence suggests that the year-to-year variability related to local weather conditions and water levels can significantly impact the severity of hypoxia and degree of recovery (Liu et al., 2014; Rucinski et al., 2016, Watson et al., 2016). Even an increased frequency of multi-day storm events during winter and spring could result in increased runoff from agricultural watersheds, such as the Maumee River Basin, and ultimately exacerbate hypoxia (Cousino et al., 2015). Another emerging issue involves the likelihood of significant primary productivity during the ice cover period with possible strong linkages to *DO* availability in the system. Relatively high under-ice phytoplankton biomass has been recorded in the central and eastern basins, comparable to or higher than values typically observed in Lake Erie in the summer (Oveisy et al., 2014; Reavie et al., 2016). This elevated winter-spring production in the central basin of Lake Erie may become available for export to other trophic levels, including bacterial decomposition at the lake bottom later in the year, when the increase in water temperature could trigger the development of hypoxia (Twiss et al., 2012; Oveisy et al., 2014). Thus, there are views in the literature asserting that the -easily biodegradable- biogenic material produced from winter-early spring algal blooms represents a considerable quantity that needs to be an integral part of any future modelling exercise that will revise the load-response curves.

4.2.4. Eastern basin Cladophora represented by dry weight biomass and stored P content: *Cladophora glomerata* is a filamentous green alga that has proliferated in the rocky nearshore zone of eastern Lake Erie since the mid-1990s, and has been responsible for the extensive fouling of the local beaches by decaying organic material (Higgins et al., 2008). Because of the meso-oligotrophic status of the offshore waters in the eastern basin, the presence of widespread *Cladophora* blooms in the nearshore area was attributed to the major reengineering of the biophysical littoral environment brought about by the invasion of dreissenids, including the profound alterations on the retention and recycling of nutrients (Tomlinson et al., 2010). Given the way this water quality problem manifests itself, the extent of beach fouling by sloughed material would have been the most suitable metric for this *ERI*. Nonetheless, the absence of a model (or an acceptable monitoring program) to evaluate the potential improvement against external nutrient loading reductions with such a metric, led to the pragmatic selection of algal dry weight biomass and stored phosphorus content as two proxies to track progress in the northeastern shoreline of Lake Erie (Scavia et al., 2016b). The achievability of a threshold value of 30 g dry weight biomass/m^2 was assessed by Valipour et al. (2016) using the high-resolution, three-dimensional *ELCOM-CAEDYM* coupled with the *EBC* model. Load–response curves were generated in an area centered on the Grand River and covering 40 km of the northern shoreline area out to the 15 m depth contour, and the predictions drawn suggested that *P* load reductions will bring about minor decline in the *Cladophora* biomass in the eastern basin (Fig. S3f). The predicted responses were influenced by the light availability, and the predicted changes were thus more noticeable at the non-light-limited shallow depths and less so at deeper depths. In fact, a small *Cladophora* biomass increase was predicted at deeper depths as a result of the improved water transparency induced by the most extreme loading reduction scenarios (Valipour et al., 2016).

Consistent with Scavia et al.'s (2016a) interpretation, we believe that there are three compelling reasons why the predicted load-response curves should be viewed with extra caution First, the domain within which *Cladophora* growth could be regulated by *SRP* concentrations is extremely low (0.2–1.0 μg P L^{-1}; see Tomlinson et al., 2010), while year-to-year variability even on the order of 1 μg P L^{-1} could result in variations of depth-integrated biomass by a factor of 3.5 (Higgins et al., 2005). Second, except from the supply by dreissenid excreta (Ozersky et al., 2009), the *SRP* nearshore concentrations are also modulated by the inflows from the Grand River as well as the nearshore–offshore exchanges. In particular, frequent upwelling events driven by favorable winds of 5–10 days period can easily increase *P* supply above saturation levels (Valipour et al., 2016). Third, although plausible explanations on the factors that accelerate the sloughing rates and their development within the *Cladophora* mats do exist, the mathematical representation of the associated processes is far from adequate. Sloughing rate is treated as a first-order loss process varying as a function of water temperature (physiological effect) and the depth of colonization (wind energy, benthic shear stress effect), but several important conceptual advancements are still overlooked (Higgins et al., 2008). For example, because of the absence of intercellular organelles or plasmodesmata, *Cladophora* filaments display limited transport of nutrient and other metabolic compounds from the surface to the base of the algal mats, and thus cellular deterioration at the base of the *Cladophora* mats can occur while cells toward the surface of the mat still actively grow (Higgins et al., 2006). Recognizing that nuisance *Cladophora* growth are controlled by both local- and/or basin-level factors, our ability to examine the potential impact of large-scale watershed management actions and draw inference with the granularity required to understand nearshore, small-scale processes is confounded by considerable uncertainty. It is unrealistic to expect mechanistic models that have been calibrated to match average conditions in the offshore waters (along with all their inherent structural and parametric uncertainties) to support predictions in the nearshore zone with accuracy <1 μg P L^{-1}. There is no evidence in the international literature that the current generation of models can go anywhere close to the required accuracy and spatiotemporal resolution required to address the *Cladophora* problem.

We thus emphasize that the local modelling efforts in Lake Erie will greatly benefit from a high-resolution monitoring of the nearshore zone to provide critical information for the existing model ensemble (Table S2). Rather than increasing the complexity of (already) over-parameterized models, the management efforts will be better supported by the development of two empirical models offering causal linkages between the abiotic conditions (e.g., *SRP*, light, temperature) in the surrounding environment and the internal *P* content and sloughing rates in *Cladophora* mats. Similar to Tomlinson et al. (2010)'s polynomial functions used to incorporate the role of temperature and light into *EBC* model, after digitizing the data originally reported by Graham et al. (1982), our proposition is to build an empirical modelling framework that will connect individual snapshots of the ambient conditions in the nearshore zone with the contemporaneous internal *P* levels at different depths during the growing season as well as the net production rates at the base of *Cladophora* mats. These two models could be used independently or in conjunction with the existing coupled *ELCOM-CAEDYM-EBC*, thereby introducing an extra layer of causality that connects the empirical characterization of microscale processes with coarser scale predictions of mathematical models (Shimoda et al., 2016b; see also following discussion).

4.2.5. Take-home Messages: In our attempt to distinguish between predictable patterns and sources of uncertainty with the load-response curves developed for the four *ERI*s, there are several lessons learned and important issues for future consideration: (*i*) the diversity of the existing modelling work in Lake Erie as well as the general evidence from the international literature suggest that the forecasting exercise related to the overall summer phytoplankton biomass in the western basin has a lot of potential to meaningfully assist the local management efforts. Our analysis highlighted the need to establish a more reliable empirical model (chlorophyll *a* versus *TP* and other potential predictors) and also improve our understanding (and subsequently their representation with the existing mechanistic models) of certain facets of phytoplankton ecology, such as the postulated degree of reliance of phytoplankton growth upon internal nutrient sources (e.g., microbially mediated regeneration, dreissenid or zooplankton excreted material in nearshore and offshore waters, respectively), the internal *P* loading from the sediments, and phytoplankton-zooplankton interactions. (*ii*) Considering the challenges with the modelling of individual phytoplankton functional groups, the forecasting exercise regarding the c*HAB* likelihood of occurrence under different loading regimes is as robust as can it be realistically expected. The coupling of empirical and process-based models offers a healthy foundation to evaluate competing hypotheses and advance our knowledge on the suite of factors that may trigger cyanobacteria dominance in Lake Erie. We just caution though that the reported range of cumulative Maumee March–July annual loads of 1679–2170 MT for achieving the c*HAB* target is likely narrow and does not fully reflect the actual uncertainty with this *ERI*. (*iii*) Our study casts doubt on the ability of the existing models to support reliable predictions regarding the likelihood to alleviate the hypoxia in the central basin of Lake Erie, given that our mechanistic understanding of sediment diagenesis, i.e., the characterization of organic matter mineralization and redox-controlled processes within different sediment layers, is still inadequate. There is a rich research agenda that should be in place with the next iteration of the adaptive management cycle, before we are in a position to predict the degree and timing of the sediment response or the likelihood of unexpected feedback loops that could delay the realization of the anticipated outcomes. (*iv*) The modelling of *Cladophora* in the eastern basin has been insightful but carries little predictive value. The proposed high-resolution monitoring of the nearshore zone and subsequent establishment of causal linkages between abiotic conditions (e.g., *SRP*, light, temperature) in the surrounding environment and the internal *P* content and sloughing rates in *Cladophora* mats are two essential steps to further augment the existing modelling work.

5 Environmental Modelling and Adaptive Management Implementation in Lake Erie

The rigorous analysis of decision problems in water quality management requires specification of ecosystem indicators that reliably reflect the prevailing conditions; an objective function to evaluate benefits and costs of alternative management strategies; predictive models formulated in terms of variables relevant to management objectives; a finite set of alternative management actions, including any conditional constraints on their use; and a monitoring program to follow system evolution and responses to management (Walters, 1986). In this regard, one of the major challenges is associated with the uncertainty in the predictions of management outcomes. This uncertainty may stem from incomplete control of management actions, errors in measurements and sampling, environmental variability, or incomplete knowledge of system behavior. Failure to recognize and account for these sources of uncertainty may lead to catastrophic environmental and economic losses. Consequently, there has been a growing interest in the policy practice of adaptive management, as it provides an iterative implementation strategy recommended to address the uncertainty associated with ecological forecasts and to minimize the impact of inefficient management plans. Adaptive implementation or "learning while doing" augments initial forecasts of management actions with post-implementation monitoring, and the resulting integration of monitoring and modelling provides the basis for revised management actions. In Lake Erie, a unique combination of statistical and mathematical models have been developed to evaluate the relationships among watershed physiography, land use patterns, and phosphorus loading, to understand ecological interactions, to elucidate the role of specific facets of the ecosystem functioning (internal loading, dreissenids), and to predict the response of the lake to external nutrient loading reductions. Consistent with the scientific process of progressive learning, the present study aimed to assist the next iteration of the modelling framework by impartially identifying strengths and weaknesses of the existing models and pinpointing essential augmentations and research/monitoring priorities in order to effectively integrate watershed and aquatic ecosystem processes (Fig. 6).

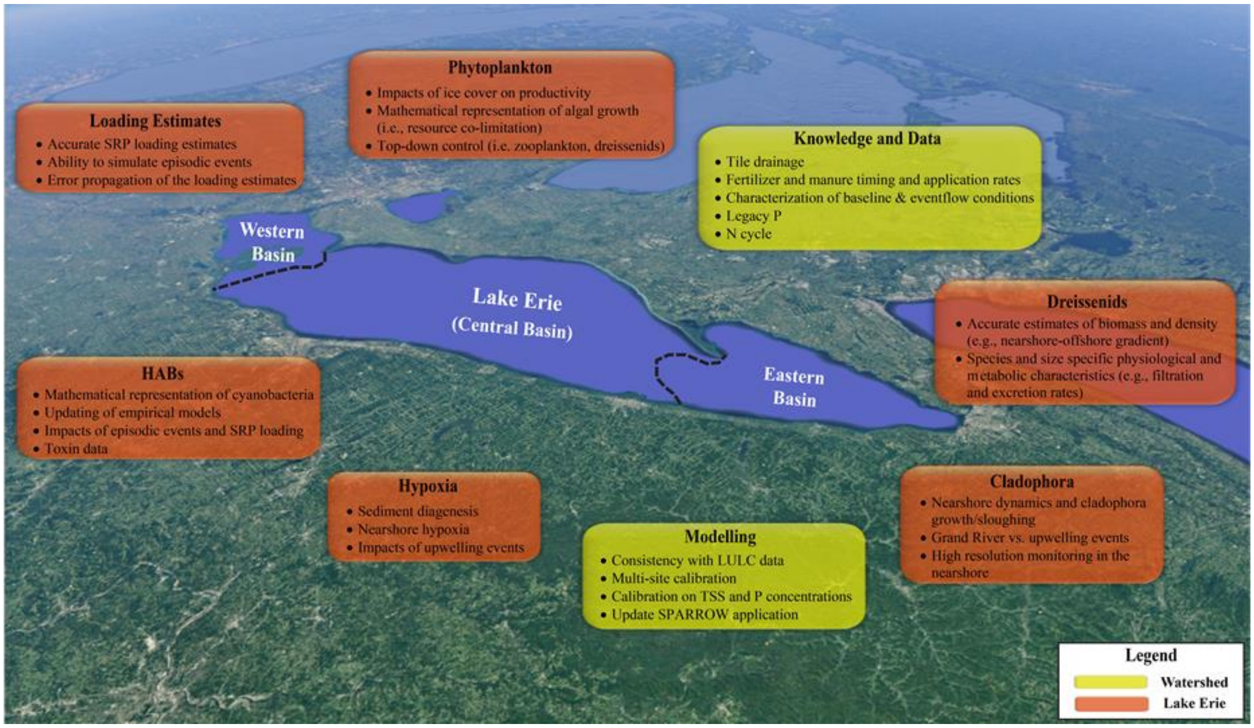


Figure 6: Knowledge gaps and sources of uncertainty to guide monitoring and improve modelling in Lake Erie (See also Table S2).

5.1 Building an integrated modelling framework in the context of adaptive management implementation: What are the recommended next steps?

Recognizing that we can never have all the empirical information to develop a completely constrained model, there is always a trade-off between knowing "much about little" or "little about much" in the environmental modelling practice. In Lake Erie, most of the existing work opted for the latter strategy, whereby complex mathematical tools have been used to advance our understanding of the mechanisms operating both in the watershed and receiving waterbody (Shimoda et al., 2018). In a recent critique, Kim et al. (2014) argued that the majority of these process-based models are profoundly over-parameterized with unproven ability to provide robust predictive statements. Regarding the latter assertion, the skill assessment results presented by Scavia et al. (2016a,b,c) were particularly favorable with respect to their ability to capture the magnitude of important eutrophication indicators, such as phosphorus loading, phytoplankton biomass, and hypoxia severity, at an aggregated spatiotemporal (seasonal/annual time scale, basin- or lake-wide) resolution. In principle, the selected coarse scales for evaluating model performance in time and space are defensible, as they are consistent with those used for the established nutrient loading targets and water quality indicators in Lake Erie (Scavia et al., 2016a,b). Nonetheless, given that the majority of these models are based on daily (or sub-daily) simulations within one-to three-dimensional spatial domains, it would seem that the bar of what constitutes an *acceptable* model performance has been lowered significantly. There are compelling reasons why this practice is problematic and should be revisited during the next iteration of the modelling framework. From a technical standpoint, evaluating model goodness-of-fit with a coarser resolution not only entails the risk to obfuscate multiple daily or location-specific errors/biases that cancel each other out when seasonally or spatially averaged, but may also detract the attention from the much-needed critical evaluation of the process characterizations derived after the calibration of (prone-to-overfitting) complex models. In particular, many of the assumptions made or parameter values assigned could be adequate to describe spatially or temporally aggregated patterns, but could also be the culprits for the misrepresentation of important aspects of the intra- or inter-annual and spatial variability (e.g., magnitude of the spring freshet, timing of algal blooms, and response of the nearshore zone to extreme precipitation events). Our independent model-fit reassessment exercise on a daily scale reinforced the importance of the latter issue by showing the distinctly inferior performance of both watershed and aquatic ecosystem models, as well as their inability to capture critical short-term or event-based facets of the simulated terrestrial and aquatic biogeochemical cycles.

In Lake Erie, the development of an ensemble of models offers the unique ability to evaluate competing hypotheses regarding the relative importance of hydrological processes and mechanisms of nutrient fate and transport within a watershed context, or the plausibility of alternative aquatic ecosystem conceptualizations; especially when complex over-parameterized models are in place with inadequate empirical information to pose any meaningful constraints. The propagation of this uncertainty through the environmental forecasts (i.e., *BMP* scenarios, load-response curves) was based merely on the generation of the uncertainty envelope from individual model predictions, without the consideration of weighting factors that consider their goodness-of-fit, bias, or model complexity (Scavia et al., 2016b, 2017). Counter to this practice, there are viewpoints in the literature advocating the development of weighting schemes to objectively synthesize ecological forecasts from multiple models (Wilks, 2002; Raftery et al., 2003; Roulston and Smith, 2003). One of the critical decisions involves the development of standards for the calibration and validation domains that will allow to rigorously evaluate the ability of a model for extrapolative tasks, i.e., forecast conditions distinctly different from those currently prevailing in the modelled system (Ramin et al., 2012a). Another criterion focuses on the goodness-of-fit of individual models as a weighting factor to determine their corresponding influence on the ensemble predictions. In particular, Ramin et al. (2012a) advocated the consideration of the performance over all model endpoints, for which observed data exist, rather than the variables more closely related to the environmental management problem at hand. In doing so, we ensure that the models included in an ensemble environmental forecast should have balanced performance over their entire structure. To put in another way, we ought to penalize the likelihood of calibration bias, whereby the maximization of the fit for a specific variable (e.g., nutrients or total phytoplankton biomass) may be accompanied by high error for other variables (e.g., individual phytoplankton functional groups or zooplankton), and thus avoid deriving forecasts founded upon models with misleadingly high weights that conceal fundamentally flawed representations of system behaviour. Other criteria for the development of ensemble weighting schemes are the consideration of penalties for model complexity that will favor parsimonious models (McDonald and Urban, 2010), and performance assessments that do not exclusively consider model endpoints but also evaluate the plausibility of the values assigned to major processes pertaining to water budget, nutrient cycles, and critical ecological pathways against empirical estimates; whenever these values exist (Arhonditsis and Brett, 2004; Wellen et al., 2015). Regarding the latter factor, it is also important to reiterate that no consistent information has been reported in the Lake Erie modelling literature regarding the relative magnitude of modelled biogeochemical fluxes, and therefore it is difficult to evaluate how "diverse" the model ensembles for the watershed or receiving waterbody are or to what extent the individual models replicate the same characterization of the system functioning.

In reviewing the credibility of the load-response curves, our study highlighted several important structural augmentations of the existing modelling tools that could increase both their heuristic and predictive values as long as commensurate empirical knowledge to constrain the mathematics becomes available from Lake Erie. If we strive to establish predictive linkages between the magnitude and timing of the response of the sediments and different loading regimes, the study of the sediment diagenesis processes is essential in understanding the control of redox chemistry on the vertical profiles of biodegradable organic matter and *P* binding forms (Gudimov et al., 2016). Albeit their conveniently simple form, the expressions of *SOD* as a function of *DO*, temperature, or even *TP* loading are primarily conceptual without adequate ground-truthing in the literature and therefore carry little predictive power. Empirical information is also needed to constrain the submodels/differential equations related to dreissenids, *Cladophora*, and zooplankton. While some progress has been made in representing the role of dreissenid mussels in the system (Verhamme et al., 2016; Karatayev et al., 2017), little work has been done to adapt the existing *Cladophora* submodel to the nearshore zone and even less so to depict the phytoplankton-zooplankton interactions in Lake Erie (Table S2). Likewise, with the shift in focus to the average conditions of the offshore waters, the nearshore zone has received less attention from the existing modelling work in Lake Erie. These areas are intermediate zones in that they can receive polluted inland waters from watersheds with significant agricultural, urban and/or industrial activities while mixing with offshore waters, having different biological and chemical characteristics. Coastal upwelling events during the early summer appear to modulate offshore-nearshore mass exchanges, whereby nutrients and hypoxic waters are injected from the hypolimnion into the nearshore Lake Erie (Valipour et al., 2016). Surface waves can also resuspend bottom sediments in the shallow waters, and as they tend to be repositories of both nutrients and contaminants, resuspension events are highly important in predicting water quality. Thus, there is a need for an integrated watershed-receiving waterbody modelling framework to shed light on the interactions of surface/subsurface hydrological inflows with in-lake hydrodynamics that shape to a large degree the dispersal of pollutants and consequently the spatial extent and magnitude of associated ecological impacts in different basin of Lake Erie (Schwab et al., 2009).

Another critical challenge revolves around the establishment of robust phytoplankton group-specific parameterizations to support predictions in a wide array of spatiotemporal domains, given the uncertainty with the derivation of distinct functional groups from fairly heterogeneous algal assemblages and our knowledge gaps of cyanobacteria ecology (Shimoda and Arhonditsis, 2016). The ability of the current generation of plankton models to reproduce succession patterns and structural shifts in phytoplankton communities has not been proven yet, and thus efforts to predict c*HAB*s with process-based models are often characterized as attempts to “run before we can walk” (Anderson, 2005). Although we do not agree with these skeptical views, we do believe that the inclusion of empirical cause-effect relationships in the model ensemble to link the nutrient loading variability with the magnitude of the summer harmful algal bloom offers a reliable complementary framework to track the anticipated response of the system (Bertani et al., 2016; Stumpf et al., 2016). Building upon the Bayesian foundation of some of these empirical tools, the next steps should involve their sequential updating as more data are acquired through monitoring, as well as the consideration of additional predictors to accommodate the role of other nutrients, light availability, water column stability, and water temperature. The latter augmentation not only will improve our predictive power, but will also allow to establish hierarchical linkages between the year-to-year variability captured by the Bertani et al.'s (2016) and Stumpf et al.'s (2016) studies and within-year conditions that ultimately lead to c*HAB* formation (Shimoda et al., 2016a).

5.2 Do we need other models to complement the *SWAT* framework in the Maumee River watershed?

In terms of the diversity of the watershed modelling framework, *SWAT* has been the only process-based model used to evaluate alternative agricultural management practices in the Maumee River watershed. Notwithstanding its conceptual and operational advantages, an important question arising is to what extent are we missing profound advancements of our understanding of watershed processes that other models can offer? To address the latter question, we compared *SWAT* against eight commonly used physically-based watershed models, i.e., the AGricultural Non-Point Source Pollution Model (*AGNPS*; Kirnak, 2002), Distributed Large Basin Runoff Model (*DLBRM*; He and DeMarchi, 2010), Dynamic Watershed Simulation Model (DWSM; Borah et al., 1999; 2002), Generalized Watershed Loading Function (*GWLF*; Borah et al., 2006), Hydrologiska Byråns Vattenbalansavdelning-Integrated Catchment (*HBV-INCA*; Crossman et al., 2013), Hydrological Simulation Program-Fortran (*HSPF*; Canale et al., 2010), *MIKE SHE* (Refsgaard and Storm, 1995), and Storm Water Management Model (*SWMM*; Rossman and Huber, 2016), regarding their strategies to capture surface runoff, groundwater, sediment transport, nutrient cycling, and channel routing (Fig. 7). For the purpose of the present study, we only provide the distinct modelling strategies that are available in the literature, but detailed description of the various models can be found in Dong et al. (2018).

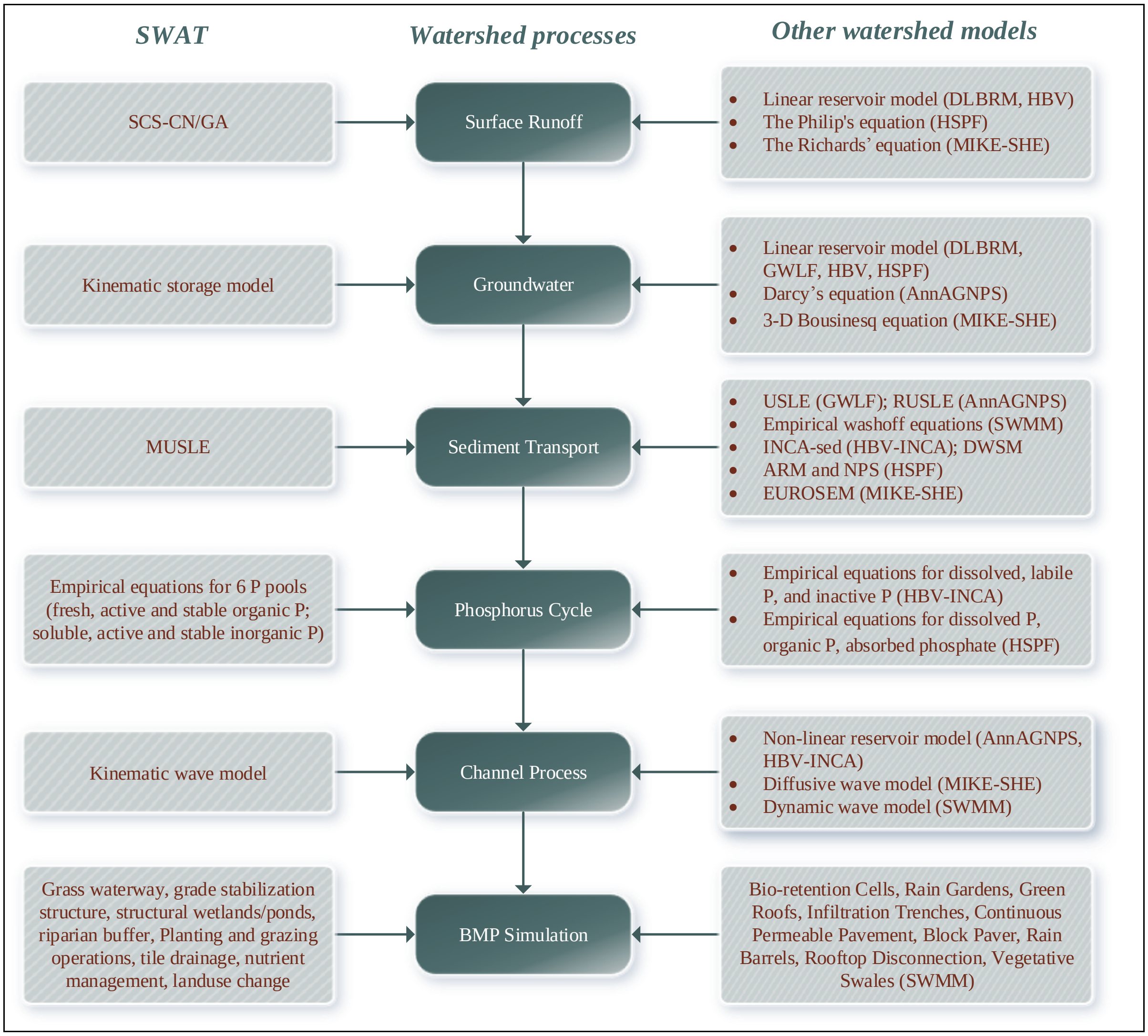


Figure 7: Comparison of *SWAT* against other process-based watershed models in terms of the representation of selected watershed processes (surface runoff, groundwater, sediment transport, phosphorus cycle, channel routing, and *BMP* simulations). The abbreviations stand for *SWAT*: Soil Water Assessment Tool; *AGNPS*: Agricultural Non-Point Source Pollution Model; *DWSM*: Distributed-Parameter Large Basin Runoff Model; *GWLF*: Generalized Watershed Loading Functions; *DWSM*: Dynamic Watershed Simulation Model; *HBV*: Hydrologiska Byråns Vattenbalansavdelning; *INCA*: Integrated Catchment Model; *HSPF*: Hydrologic Simulation Program *FORTRAN*; *SWMM*: The Storm Water Management Model; *SCS-CN*: The Soil Conservation Service Curve Number method; *GA*: Green-Ampt infiltration method; *MUSLE*: Modified Universal Soil Loss Equation.

To quantify the potential magnitude of surface runoff, *SWAT* uses the empirical Soil Conservation Service Curve Number (*SCS-CN*) method based on the antecedent moisture condition and the hydrologic soil group of a particular location at a given day. The *SCS-CN* method lumps rainfall interception, depression storage, and soil infiltration as the initial abstractions, which are assumed to account for 20% of the potential maximum retention. The surface runoff is estimated by subtracting that amount from the rainfall total volume, and therefore the total precipitation volume should exceed the initial abstraction before any runoff is generated. Because of the limitations of the *SCS-CN* method in estimating cumulative runoff depth and peak flows (Borah et al., 2007), *SWAT* also provides a sub-daily simplified model, the Green-Ampt (*GA*) infiltration method, by postulating a uniform movement of water from the surface down through the deep soil with a sharp wetting front. The *HBV* and *DLBRM* models used the linear-reservoir concept to represent the rainfall-runoff process at the watershed scale, based on the assumption that the overland flow is linearly correlated with the water storage. *HSPF* uses the Philip equation, which is a sub-daily method simplified from the Richards' equation (Richards, 1931). The latter approach is considered by the *MIKE SHE* model and describes the vertical movement of water through the soil profile using Darcy's law. The flow rate through porous media is proportional to the hydraulic conductivity, which in turn is estimated from the soil water content (Baver et al., 1972). Compared to other models, the Richards' equation could more accurately quantify vertical water percolation and dynamic unsaturated flow based on various soil properties.

In terms of the representation of subsurface process, *SWAT* incorporates a kinematic storage model to simulate lateral flow in the unsaturated zone. The saturated zone is conceptualized as an unconfined shallow aquifer and confined deep aquifer. Groundwater flowing into the main channel is assumed to be linearly correlated with hydraulic conductivity and the changes of the water table height. By contrast, a simple linear reservoir model is used in *DLBRM*, *HBV*, *HSPF*, and *GWLF*, whereby interflow and groundwater are linearly proportional to moisture content of the unsaturated zone and saturated zone, respectively. Conceptually similar with the division of shallow and deep aquifers in *SWAT*, *HSPF* divides the saturated zone into two storage reservoirs: active and inactive groundwater. *AnnAGNPS* also uses a simple empirical method, the Darcy's equation, assuming that lateral flow is linearly related to saturated hydraulic conductivity and hydraulic gradient. Compared to subsurface modelling in *SWAT*, both the linear reservoir model and the Darcy's equation are simpler representations of the subsurface system. Conversely, *MIKE SHE* provides a more reliable physically-based *3D* saturated zone model to explicitly represent the vertical and spatial characteristics of the subsurface profile. *MIKE SHE* incorporates a *3D* Finite Difference Method numerical engine, which is theoretically similar to *MODFLOW* (MODular 3D Finite-Difference Ground-Water FLOW Model) (McDonald and Harbaugh, 2003). The subsurface modelling in *SWAT* could be complemented with *MIKE SHE* by tracking the vertical solute transport and providing a comprehensive, fully dynamic depiction of the hydrological interplay between surface and subsurface layers.

One of the commonly used strategies to quantify soil erosion, the modified Universal Soil Loss Equation (*MUSLE*), is used in *SWAT*. *MUSLE* is developed from the original *USLE* method by replacing the rainfall-erosivity (*R*) factor with a runoff energy factor. The Revised *USLE* (*RUSLE*) used in *AnnAGNPS* is another modification by converting the soil erodibility (*K*) factor to a time-varying parameter. Counter to these *USLE*-based empirical approaches, *DWSM*, *HBV-INCA*, *HSPF*, and *MIKE-SHE* offer physically-based methods that explicitly accommodate the physical detachment/reattachment and transport processes. The European Soil Erosion Model (*EUROSEM*) used in *MIKE SHE* considers soil detachment affected by both raindrop and leaf drainage, which enables explicit representation of effects of vegetation heights. By contrast, *HSPF*, *INCA* and *DWSM* erosion models ignore the effects of leaf drip. Thus, the sediment erosion module in *MIKE SHE* introduces an advanced representation of our contemporary understanding of the soil erosion process, which could be used to complement the empirical *MUSLE* method in *SWAT*. Soil phosphorus (*P*) simulation with *SWAT* is based on the *EPIC* (Erosion Productivity Impact Calculator) model (Sharpley and Williams, 1990), which includes three organic *P* pools (fresh, active and stable) and three inorganic pools (solution, active and stable). In *SWAT*, fertilizer, manure, and residues are the input sources of soil *P*, which could be removed by plant uptake, water flow and soil erosion. The plant *P* uptake is dependent on plant growth, simulated as a function of the leaf area growth, light interception, and biomass production. One of the strengths of the *EPIC* model used in *SWAT* is the explicit simulation of the daily plant growth, whereas other models, like *INCA* and *HSPF*, represent plant growth either through a seasonal plant growth index, or a simple empirical first-order kinetics equation. *SWAT* also takes into account the residue decay and mineralization, which is not considered with *HSPF* and *INCA* and may thus limit their ability to evaluate the importance of legacy *P*. Nevertheless, a major limitation of *SWAT* in simulating long-term soil *P* dynamics is the assumption of a constant equilibrium adsorption/desorption concentration. On the contrary, *HSPF* adjusts the adsorption rates by soil temperature, and furthermore, *INCA* varies the equilibrium inorganic *P* concentration based on the mass of *P* in the labile soil store. The latter feature can be quite critical when evaluating the long-term watershed responses to various agricultural management strategies.

Existing submodels of water routing in channels could be divided into four main categories spanning a wide range of complexity: the dynamic wave model; the diffusive wave model; the kinematic wave model; and the non-linear reservoir model. The dynamic wave model used in *MIKE SHE* and *SWMM* is the most complex physically-based approach based on the continuity and momentum equations (Tayfur et al., 1993). The diffusive wave model uses simplified momentum equations by downplaying the role of local and convective acceleration. *SWAT* uses the kinematic wave model, which incorporates the most simplified momentum equation by omitting the pressure gradient and acceleration terms (Miller, 1984). The non-linear reservoir model divides the channel segments into a series of reservoirs with uniform water surface. Since *SWAT* omits pressurized flow and backwater effects, it is not capable of simulating pipe flows, whereas a fully dynamic wave equation to model water routing would be applicable to both open channels and closed pipes.

Considering the different strengths of the watershed models reviewed, *SWAT* could be complemented by the modules of other watershed models, especially for surface runoff, groundwater and sediment erosion processes. For the hydrological and sediment processes, *MIKE SHE* seems to be more up-to-date with respect to the mechanisms considered, assuming that local empirical knowledge is available to constrain the additional parameters. Although *SWAT* has the advantage to explicitly simulate the daily plant growth, it could be improved by adopting the dynamic *P* equilibrium concentration. *MIKE SHE* and *SWMM* are superior to *SWAT* in channel routing because of their capability to simulate pipe flows. *SWAT* is also more suitable for agricultural *BMP*s (e.g., terracing, contouring, strip cropping, tillage operations, crop rotations, and fertilizer application), while the urban *BMP* modules in *SWMM* (e.g., rain gardens, green roofs, infiltration trenches, permeable pavement, and vegetative swales) offer a more reliable alternative (Dong et al., 2018). Another challenge with the existing *SWAT* applications in the Maumee River watershed was the consistent event-flow underestimation, which can be potentially ameliorated by the use of an alternative runoff estimator, such as a Green-Ampt method, instead of the conventional *CN* method (James et al., 1992; King et al., 1999). Nonetheless, it is important to recognize that the characterization of surface runoff and subsurface processes during flow events is largely unknown in the area, and therefore the design of high frequency, event-based, water quality sampling coupled with water stable isotope analysis (^{18}O and ^{2}H) should be one of the priorities in our efforts to rectify the misrepresentation of extreme flow conditions (Klaus and McDonnell, 2013; Kim et al., 2018). In the same context, recent advancements in hydrology suggest that baseline conditions and extreme events may be associated with distinct flow mechanisms, and thus two major strategies have been proposed to accommodate threshold behavior in watershed models (Zehe and Sivapalan, 2009). The first strategy is the introduction of a two-domain conceptualization of soil water movement into numerical watershed models, whereby flow through the soil matrix or macropore flow is responsible for small and large runoff events, respectively (Zehe et al., 2001). The second approach postulates that the watershed operates in multiple states or modes of behavior, and the identification of which can be explicitly accommodated through the model calibration process (Ali et al., 2013). A characteristic example of this strategy is the Bayesian hierarchical framework used to calibrate the *SWAT* model in Hamilton Harbour (Ontario, Canada), which enabled the identification of precipitation thresholds that trigger shifts to alternative watershed states as well as state-specific parameters to depict extreme states with higher propensity for runoff generation (Wellen et al., 2014a,b).

Together with the process-based modelling work in the Maumee River watershed, it is also critical to have simpler empirical models in place that not only provide predictive statements confined within the bounds of data-based parameter estimation, but also to constrain processes/fluxes parameterized by mechanistic models or even to validate the corresponding forecasts drawn. A characteristic example of the potential benefits of a data-driven model is the use of *SPARROW* to validate the predicted spatial distribution of phosphorus loads in the Maumee River watershed by the five *SWAT* applications (Scavia et al., 2016c). *SPARROW* is a hybrid mechanistic-statistical model, with empirically-based parameters (i.e., land-to-water delivery coefficients, nutrient export from different land uses, in-stream attenuation rates, reservoir settling velocities) used to estimate nutrient loading from a series of hydrologically linked catchments and thus to delineate areas of high risk in many Great Lakes watersheds (Wellen et al., 2014a,b; Kim et al., 2017). Interestingly, the *SPARROW* spatial projections were not in agreement with the *SWAT*-ensemble predictions of nutrient-export hot-spots in the northwestern part of the Maumee River watershed (Fig. S1), and instead were suggestive of an extensive area in the southern/ southwestern Maumee River watershed with *TP* loading estimates distinctly higher than 150 kg km^{-2} yr^{-1} (Scavia et al., 2016c). In the same context, Kim et al. (2018a) used the Base Flow Index (*BFI*) map as an independent source of information to reconcile these projected discrepancies in the Maumee River watershed attributes (Fig. S4). *BFI* is a measure of the ratio of long-term baseflow to total stream flow, representing the slow continuous contribution of groundwater to river flow, and therefore low (high) *BFI* values suggests higher (lower) likelihood of surface runoff which in turn can lead to higher (lower) suspended solid and particulate phosphorus loads. The consistently lower values of the empirically obtained baseflow index in the southern/ southwestern Maumee River watershed appear to be closer to the *SPARROW* rather than *SWAT* spatial predictions (Fig. S4). While the latter result may not be an evidence for unequivocally ground-truthing either of the two models, it does highlight the aforementioned need for the current members of the *SWAT*-ensemble to be recalibrated against data from multiple sites across the watershed, as well as to revisit some of the fundamental assumptions regarding the fertilizer/ manure application rates in the croplands or the spatial drainage of soils.

The existing *SPARROW* application in Lake Erie is part of the regional model developed for the Upper Midwest (Great Lakes and Upper Mississippi, Ohio, and Red River Basins or *MRB3* model as referred to in Robertson and Saad, 2011), which comprised 810 sites of TP empirical loading estimates. Lake Erie accounted for less than 6% of those sites, while the median and 90% percentile of the local loading estimates were 30% and 70% lower than the corresponding values in the entire dataset, respectively (Material S3 in Supporting Information section of Robertson and Saad, 2011). The use of "cross-sectional" datasets over broader regions has been one of the pillars of the *SPARROW* modelling enterprise, as this practice is deemed more suitable in unravelling the complex patterns in a watershed context while the significant loading range across the calibration locations typically leads to well-identified parameters and a larger application domain (Alexander et al., 2004). To mitigate the impact of potential outliers or non-representative calibration data, *USGS* has implemented a number of procedures, including subjective data censorship of daily records for both flow and nutrient concentrations in case they are obstructing satisfactory Fluxmaster regression fit (Schwarz et al., 2006); water quality monitoring stations with number of records below a certain number are filtered out from further consideration; and stations with Fluxmaster model errors >50% are omitted from *SPARROW* calibration datasets (Robertson and Saad, 2013; Neumann et al., 2018). Nonetheless, these practices can neither address the problem of datasets that disproportionately consist of baseline- rather than event-flow samples (Richards et al., 2013; Long et al., 2014; 2015), nor do they overcome the fact that the assumption of regionally common parameters could introduce watershed-specific bias in the characterization of fundamental processes, such as nutrient export from different land-uses, land-to-water delivery, and in-stream attenuation rates. An appealing alternative that could rectify many of these problems will be the development of a Great Lakes *SPARROW* model that narrows the focus of the original *MRB3* model, while maintaining its "global" character (Fig. 8). Importantly, the rigid common parameter estimates over the entire spatial model domain can be relaxed by the use of a hierarchical structure that allows to estimate watershed-specific parameters, and thus accommodate the spatial variability within the Great Lakes basin. In addition, rather than the strict data censoring currently implemented, the *SPARROW* practice should become more inclusive and instead the calibration datasets could be coupled with measurement-error models to characterize our degree of confidence on their quality or to accommodate the serial correlation among nested subwatersheds (Carroll et al., 2006; Balin et al., 2010; Wellen et al., 2012; 2014; Kim et al., 2017). This is an important exercise that will consolidate the presence of an empirical modelling tool to guide the delineation of nutrient hot-spots alongside the process-based modelling work.

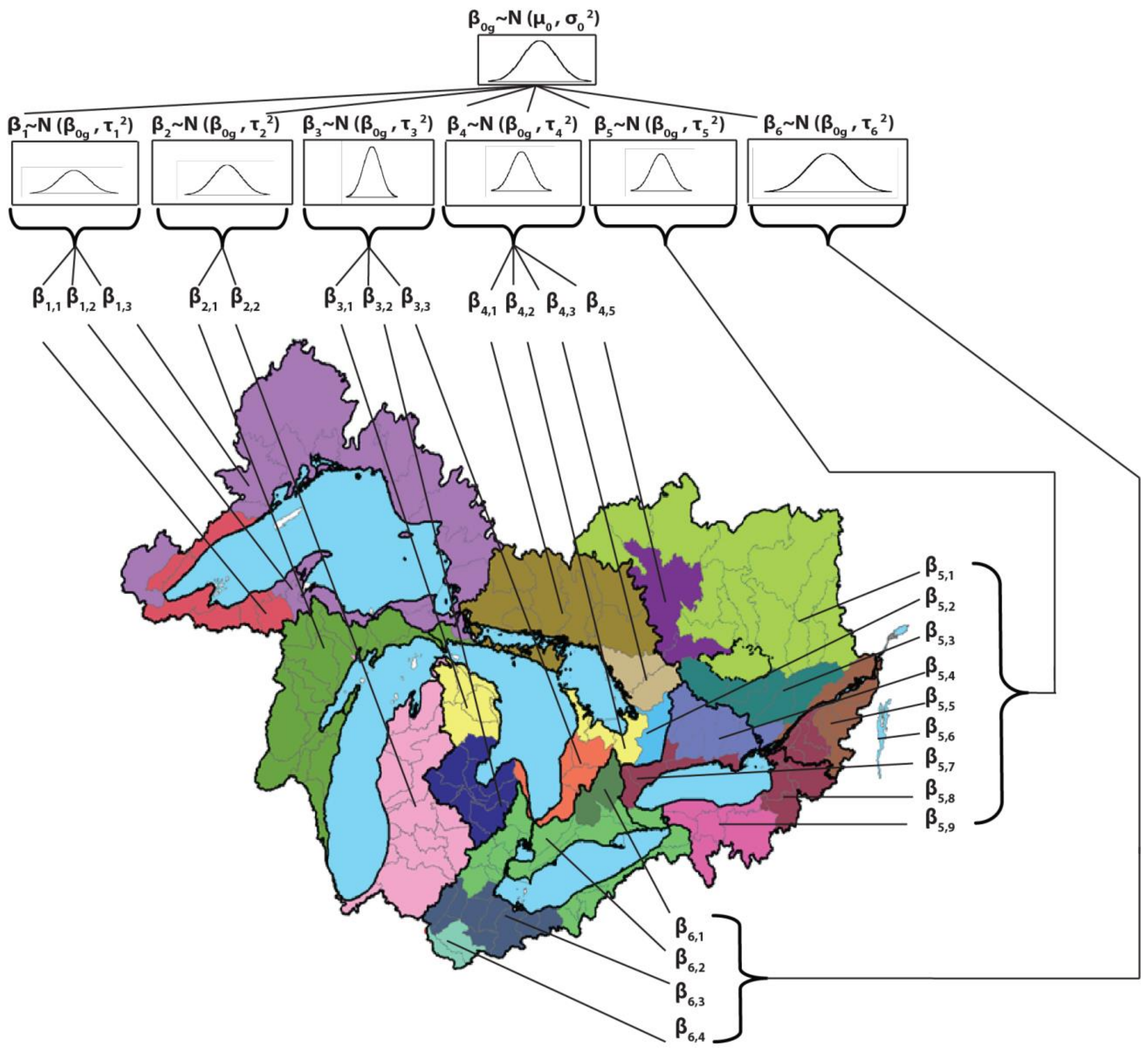


Figure 8: The case of a Great Lakes hierarchical *SPARROW*: Coupled with the mechanistic tools for the Maumee River watershed, empirical (*SPARROW*-like) models geared to depict basin-specific (rather than continental or regional) nutrient loading conditions can offer a multitude of complementary benefits, such as validate the spatial delineation of hot-spots with higher propensity for nutrient export, narrow down the uncertainty of processes/fluxes parameterized by mechanistic models, and obtain predictive statements constrained within the bounds of data-based parameter estimation.

5.3 Risks and uncertainties with the implementation of Best Management Practices: What does the literature suggest?

A variety of costly *BMP*s have been designed to mitigate pollution from diffuse sources in agricultural and urban areas (Sharpley, 2006; Dietz, 2007; Edwards et al., 2016; Leitão et al., 2018). Although their implementation has been based on the stipulation that both their short- and long-term effectiveness are guaranteed, emerging evidence is suggestive of moderate water quality improvements in many watersheds and broad variability in their performances, often much lower compared to the specs of the original design from *BMP* experimental studies (Kleinman et al., 2011; Jarvie et al., 2013). This form of scenario uncertainty can be attributed to a number of factors, such as suboptimal design, lack of landowner participation (Figs. S5a,b), erroneous selection of *BMPs* (Fig. S5c), failure to address non-point pollution sources, inadequate coverage of the watershed, lag time between *BMP* implementation and distinct improvements of downstream conditions, different efficiency between particulate and soluble forms (Figs. S5d,e), and variability induced by extreme events and other weather-related anomalies (Meals et al., 2010; Liu et al., 2017). In Lake Erie, Smith et al. (2017) noted that the majority of local farmers apply *P* fertilizers at or below the current recommendations, and thus asserted that the main culprit for the recent re-eutrophication could be the lack of appropriate fertility guidance and practices to protect water quality. The same study also questioned whether the “law of unintended consequences” has received sufficient consideration in the local decision-making process, as environmental interventions can conceivably have long-term damaging effects on ecosystem services given our limited knowledge of complex system interactions (May and Spears, 2012; Smith et al., 2017).

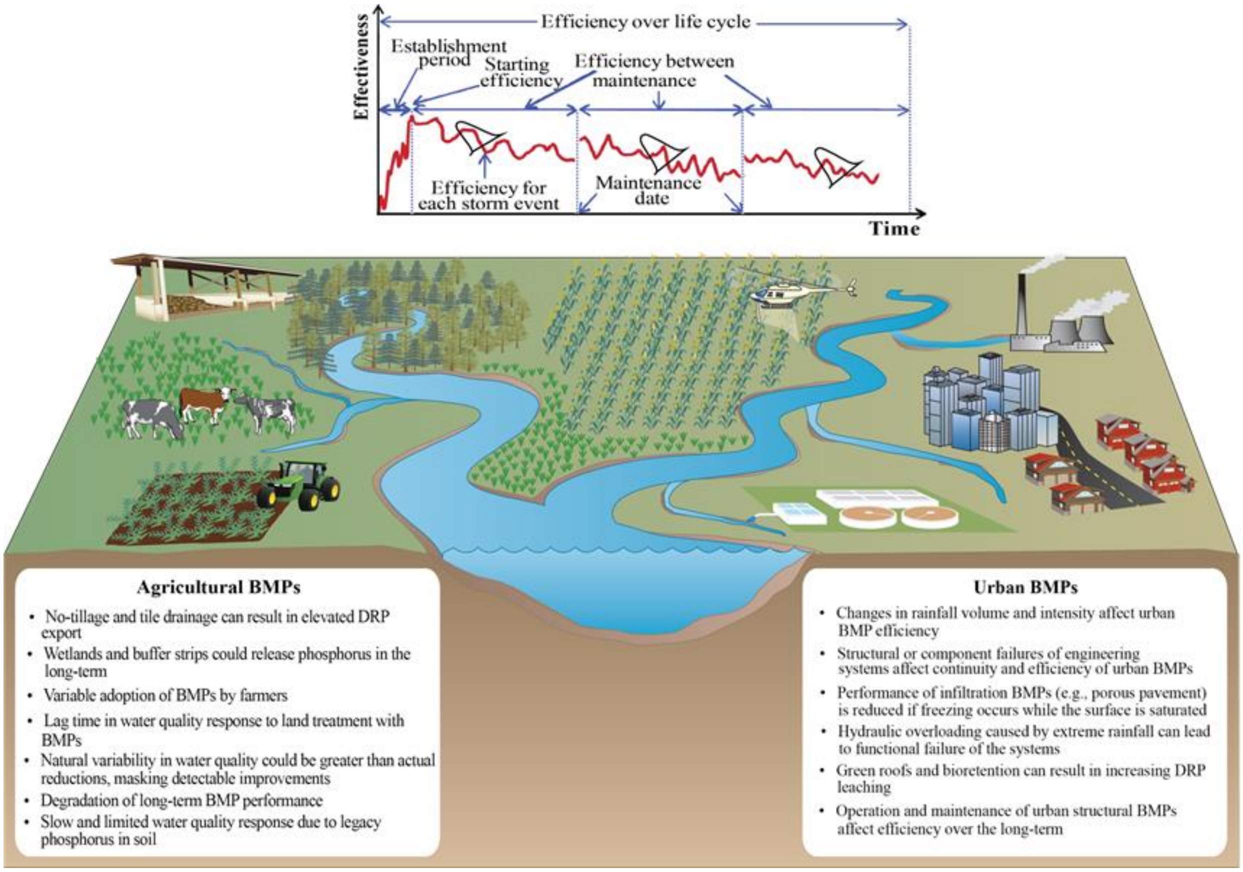


Figure 9: Risks and uncertainties with the *BMP* implementation of Best Management Practices in the Maumee River watershed. Our study highlights the importance to design land-use management scenarios that accommodate recent conceptual and technical advancements of the life-cycle effectiveness of various *BMP*s, the variability in their starting operational efficiency, and differential response to storm events or seasonality.

In the same context, Osmond et al. (2012) raised concerns that many important empirical findings from past conservation practices across North America have not been incorporated into current *BMP* guides. For example, earlier work in the area cautioned that the focus on sediment erosion control (no-till conservation, buffer strips, and fall fertilization) may entail a trade-off effect with elevated losses of bioavailable phosphorus (Logan, 1979; Gebhardt et al., 1985) and indeed recent studies by Jarvie et al. (2017) and Baker et al (2017) have attributed the re-appearance of *HAB*s to the unintended consequences from conservation decisions adopted 20-50 years ago (Fig. 9). More recently, Liu et al. (2017, 2018) identified that *BMP* performance assessments are predominantly based on short-term experimental studies, whereas long-term monitoring has registered variable performance trends. For example, Mitsch et al. (2012) has observed a gradual degradation of constructed wetlands effectiveness for *SRP* removal within 15 years of monitoring, while Kieta et al. (2018) reported limited efficiency of vegetative buffer strips (*VBS*) in Great Lakes basin, where the majority of nutrients are transported with spring freshet during the non-growing season. Similarly, Li and Babcock (2014) reported long-term orthophosphate areal export rates from green roofs comparable to those of highly intensive agricultural areas. In order to minimize the discrepancy between expected and actual environmental effects, Liu (2018) proposed a framework to incorporate *BMP* life-cycle effectiveness into watershed management plans by explicit accounting for: (*i*) the variability in starting efficiency of each *BMP* type to reduce the severity of runoff and pollutant concentrations due to local condition differences and installation practices; (*ii*) an intrinsic variability of operational performance due to watershed geophysical conditions, differential response to storm events, and seasonality; (*iii*) a non-linearity of *BMP* effectiveness in response to different loading regimes as well as an expected decline in performance over time, which in turn enforces the need for regular maintenance; and (*iv*) a lagged manifestation of water quality improvements after *BMP* adoption due to nutrient spiraling downstream or recycling in receiving water bodies (Fig. 9).

Promoting watershed management plans often requires financial incentives, such as tax credits, cost-sharing, reimbursements, insurance and certification price premiums (Tuholske and Kilbert, 2015). The aforementioned discrepancy in timing between *BMP* implementation and water quality improvement can make the financial incentives unappealing, if we opt for the "pay-per-performance" practice. Failure of selected *BMP*s to achieve loading reduction targets should be viewed cumulatively as direct budget losses, environmental capital depreciation, and socio-economic values at risk (Farber et al., 2002; Wolf and Klaiber, 2017). The consideration of *BMP* uncertainties into scenario analysis would introduce financial risk assessment in strategic agro-environmental management decisions by weighting the amount of the proposed financial incentives with non-attainment risks of nutrient reduction goals (Palm-Foster et al., 2016). The Chesapeake Bay *BMP* review protocol can serve as an exemplary case of comprehensive validation guidance of *BMP* effectiveness based on rigorous assessment of both treatment risks (known probabilities associated with *BMP* performance) and uncertainty (lack of knowledge surrounding these probabilities). The *CBP* protocol is based on transparency and inclusivity, and as such it considers detailed literature review, expert elicitation, data collection from local *BMP*s, and rigorous analysis (CBP, 2015).

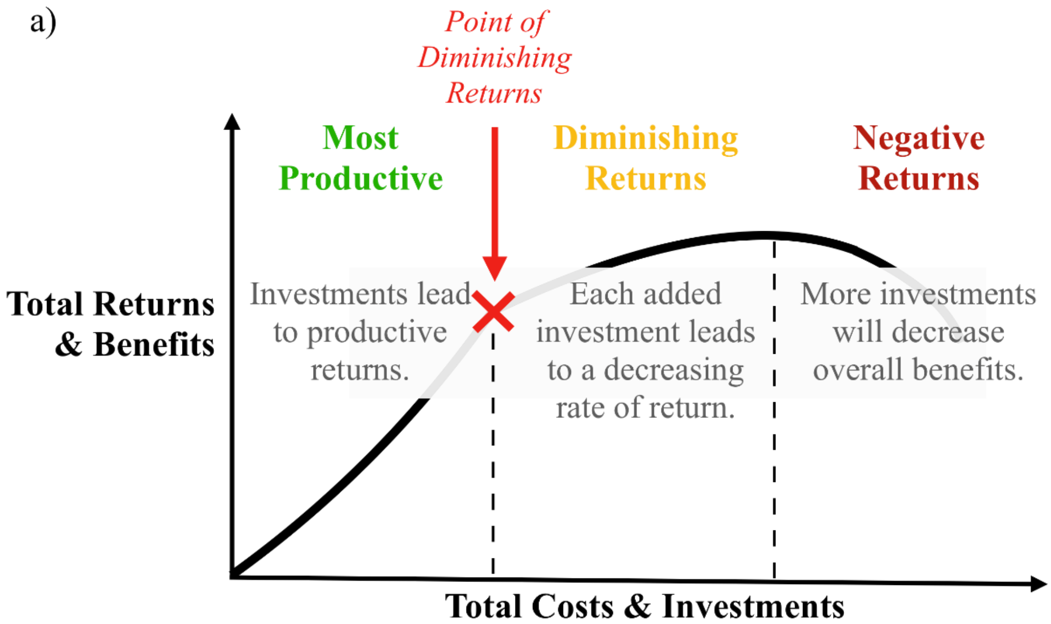
To the best of our knowledge, none of the current watershed models accounts for the life-cycle non-stationarity or overall uncertainty in *BMP* effectiveness. In particular, *SWWM5* does consider concentration-dependent removal of pollutants with specific *BMP*s during peak and base flows, but still relies on deterministic values of statistically significant median influent- and effluent-event concentrations (Rossman and Huber, 2016). Other major ecohydrological models, such as *SWAT* and *HSPF*, are either based on a deterministic (pre-specified constant) nutrient removal effectiveness or on empirical relationships of variable statistical power (Dorioz et al., 2006). A characteristic example is the *SWAT* model which considers the impact of vegetative filter strips (*VFS*) on dissolved phosphorus removal as a linear function of surface runoff reduction. Nonetheless, the corresponding regression model explains less than 30% of the observed variability, while the empirical reduction efficiency ranges from 43% to -31% near zero runoff reduction (Dillaha et al., 1989). As a first step to accommodate *BMP* uncertainty, we thus propose a moderate enhancement with a stochastic time-invariant representation of *BMP* effectiveness in watershed models (Griffin, 1995), followed by the introduction of time-variant probability distributions for *BMP* life-cycle performance (Liu et al., 2018). The proposed stochastic augmentation would allow sampling over the uncertainty of *BMP* scenarios with Monte Carlo simulations, thereby providing a pragmatic tool to assess the likelihood of the achievability of the proposed nutrient loading reduction goals. These probabilities can then be subjected to sequential updating through the iterative monitoring-modelling-assessment cycles of adaptive management, whereby our degree of confidence on the success of a selected *BMP* strategy can be refined.

5.4 Integration of ecosystem services with the Lake Erie modelling framework: An optional augmentation or an emerging imperative?

Ecosystem services are the benefits that humans directly or indirectly gain from ecosystem functions (Costanza et al., 1997). Viewing ecosystems as providers of economically valuable benefits to humans, the concept of ecosystem services effectively links their structural and functional integrity with human welfare. Lake Erie, in particular, provides numerous valuable benefits by supplying drinking water for over 11 million people, supporting a $50 billion industrial sector that encompasses tourism, boating, shipping, and fisheries, providing over 240,000 jobs in both the American and Canadian economies, and offering habitat for ecologically, culturally, and economically important biotic communities (Lake Erie Improvement Association, 2012). There is, however, a pressing need to collectively protect ecosystem services that are at risk in the current degraded state of Lake Erie. Given that environmental policy affects both the ecosystem state and the provision of services that human societies benefit from, we argue that the efficacy of the local restoration efforts will be significantly enhanced by a rigorous framework that quantifies the economic benefits from a well-functioning ecosystem. Rather than solely acknowledging their vulnerabilities, the actual quantification of the value of ecosystem services is critical when considering trade-offs among diverse policy decisions.

The rationale behind ecosystem valuation is to explicitly describe how human decisions affect ecosystem service values and to express these changes in monetary units that allow for their incorporation in the decision-making process (Pascual et al., 2010). Current markets provide information about the value of a limited subset of ecosystem services that are priced as commodities (Pascual et al., 2010), which poses challenges in our ability to estimate values of a comprehensive set of ecosystem services typically considered in the decision-making process (Millennium Ecosystem Assessment, 2005). To best communicate the trade-offs among policy choices, ecosystem service valuation must examine the marginal improvement in ecosystem services attributable to a policy change. For example, Isely et al. (2018) estimated that a $10 million investment to restore the Muskegon Lake Area of Concern would have a return on investment of approximately 6:1 and an added $50 million in environmental value over a 20-year period due to increased property values and a more attractive recreational environment. Although extensive resources and capital are required to conduct ecosystem service valuation, the outcome of such an exercise places a premium on the communication of policy trade-offs in economic terms (commercial goods/services or non-market values such as the average consumer's willingness to pay), thereby increasing stakeholder engagement and societal relevance of conservation actions (Egoh et al., 2007).

a)



b)

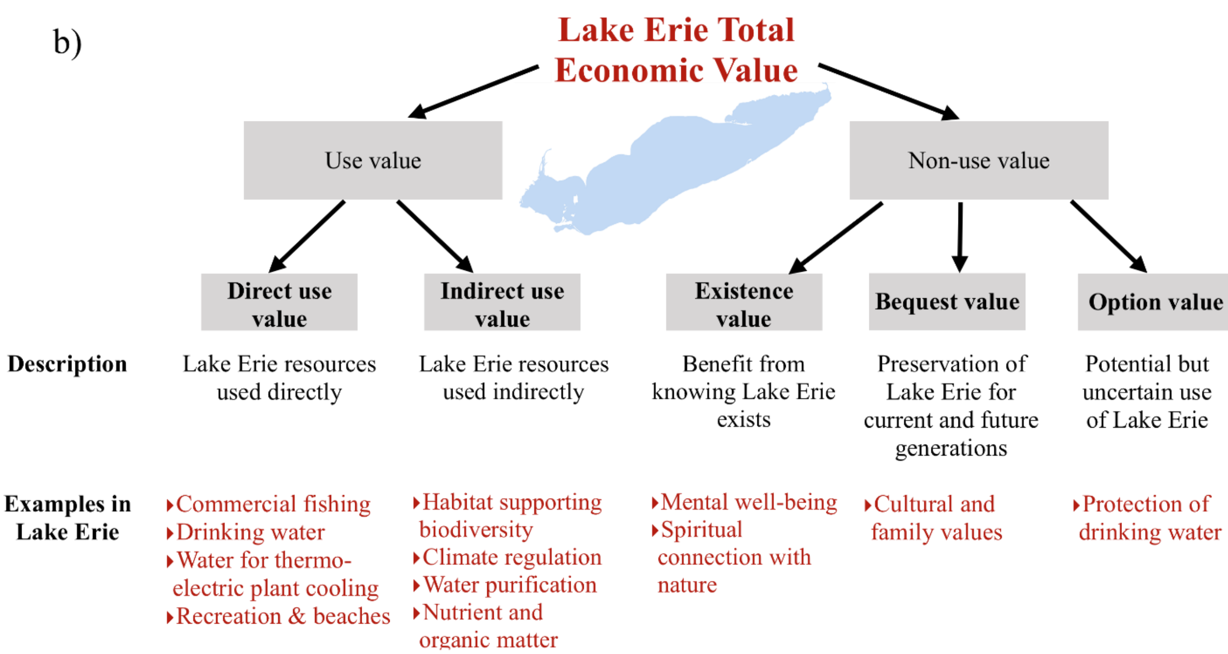


Figure 10: (*a*) Benefits accrued resulting from costs invested in an environmental restoration project. At the beginning of each restoration effort, the total returns and benefits are typically commensurate with the costs and investments, but this pattern may not hold true after a certain point, where we get diminishing (and ultimately negative) returns and marginal benefits. (*b*) Breakdown of Lake Erie's ecosystem services using the Total Economic Value framework.

Economic values of ecosystem services can help policy-makers determine the optimal degree of investment and action needed at each time step by defining the monetary trade-offs from different courses of management action (Fig. 10a). At the beginning of each restoration effort, the total returns and benefits are typically commensurate with the costs and investments, but this pattern may not hold true after a certain point, where we get diminishing (and ultimately negative) returns and marginal benefits. Viewed it from this perspective, the question arising is how likely is to experience environmental improvements proportional to the socioeconomic investments required (steep linear segment in Fig. 10a), given the presence of a wide array of feedback loops, ecological unknowns, and external factors (i.e., internal loading, dreissenid mussels, different trends between *TP* and *DRP* loading, changing climate and increased frequency of extreme events) in Lake Erie? Even more so, our analysis also highlighted an additional layer of complexity that we need to factor in during the decision making process; namely, as we opt for drastic (and likely more costly) management actions that differ significantly from the current conditions (right end of Fig. 10a), the forecasting error increases significantly (Figs. S2c,d) and so does the likelihood of realizing benefits that are distinctly lower than our original investments. Do we have enough leeway to keep the investments to the environment going? While these assertions seem to paint a pessimistic picture about the challenges and associated risks with the next steps in Lake Erie, it is important to delve into (somewhat underappreciated) ideas, such as the total economic value (*TEV*) of an ecosystem, the degree of our knowledge of the monetary value of ecosystem services in Lake Erie, and the mismatch between the scales where environmental goals are being set and the spatiotemporal domain that predominantly influences the perception of the public (Ramin et al., 2018; Kim et al., 2018b).

To facilitate ecosystem service valuation, the total economic value framework can relate a wide array of ecosystem services to human well-being in monetary terms (Fig. 10b). Direct use values are derived from the uses made of Lake Erie's resources and services, such as drinking water and the natural environment for recreation, while indirect use values are associated with Lake Erie's natural functions, such as nutrient removal and ability to provide fish habitat, refugia, and nursery (Gilpin, 2000). *TEV* typology also helps to identify non-use values that are unrelated to present or future uses, but instead reflect the value associated with the Lake Erie's existence: option, bequest, and existence values (Gilpin, 2000). Option value is the willingness to pay a certain amount today to ensure the available use of a benefit provided by Lake Erie in the future. Bequest value refers to the willingness-to-pay to preserve Lake Erie for the benefit of other people, both in the present and future. Existence value is the value attached to knowing that Lake Erie and its benefits exist, even if the individual does not intend to ever actively use them. The Great Lakes, including Lake Erie, provide a wide array of ecosystem services, although they have yet to be comprehensively inventoried (Steinman et al., 2017). Efforts have been underway to rigorously assess the status of ecosystem services and facilitate future valuation studies in Lake Erie. Allan et al. (2017) mapped the distribution of ecosystem services in the Lake Erie basin, while Annis et al. (2017) delineated optimal areas for the conservation of multiple ecosystem services in the nearshore zone of western Lake Erie.

In this context, research on ecosystem service valuation in Lake Erie has concentrated in water quality improvements, erosion risk reduction, recreation, and recreational fishing (see Table S3 for details on the methods used for these valuation studies). Brox et al. (1996) conducted a contingent valuation survey to estimate a willingness to pay of $4.50 per household per month (19% of the average water bill) for residential water quality improvements. Likewise, Kriesel et al. (1993) found that the closer a lakefront property was to Lake Erie, the more the homeowner was willing to pay for to reduce risk of damage from shoreline erosion. Building upon this finding, Dorfman et al. (1996) predicted that owners of high-risk properties would pay an average of $37,826 to effectively eliminate erosion risk, a fairly high amount given that the average selling property price in the study sample was $127,800 at that time. To estimate the value of reducing beach advisories in Lake Erie, Murray and Sohngen (2001) surveyed visitors at 15 Lake Erie beaches in the summer of 1998 and estimated the average seasonal benefits of reducing one advisory to be $28 per visitor per year, while Chen (2013) recently projected that day trips to a public Great Lakes beach (including Lake Erie) was valued at $32-39 per visitor. Importantly, Palm-Forster et al. (2016) estimated that a full-season closure for a single public beach in Lake Erie would result anywhere from $1.96 to $2.21 million depending on the valuation method used. Along the same line of evidence, Hayder and Beauchamp (2014) estimated that in 2018, the Great Lakes (including Lake Erie) provided approximately $7.76 billion in recreational benefits (recreational boating, wildlife viewing, and beach and lakefront use) and that value would increase to $354 billion in 2068. In the same context, Kelch et al. (2006) found that angling in Lake Erie was valued at $36-46 per trip and also showed that an annual $0.6 million stocking program could result in a river steelhead fishery of $12-$15 million per year in Ohio. Sohngen et al. (2015) estimated that angling trips in Lake Erie could be valued up to $88 per trip or $67.1 million per year, if we also consider the value of fish catches, and Wolf et al. (2017) estimated that a wide-scale, summer-long algal bloom in Lake Erie would reduce fishing licenses issued by 3,600 and fishing expenditures by $2.25-5.58 million.

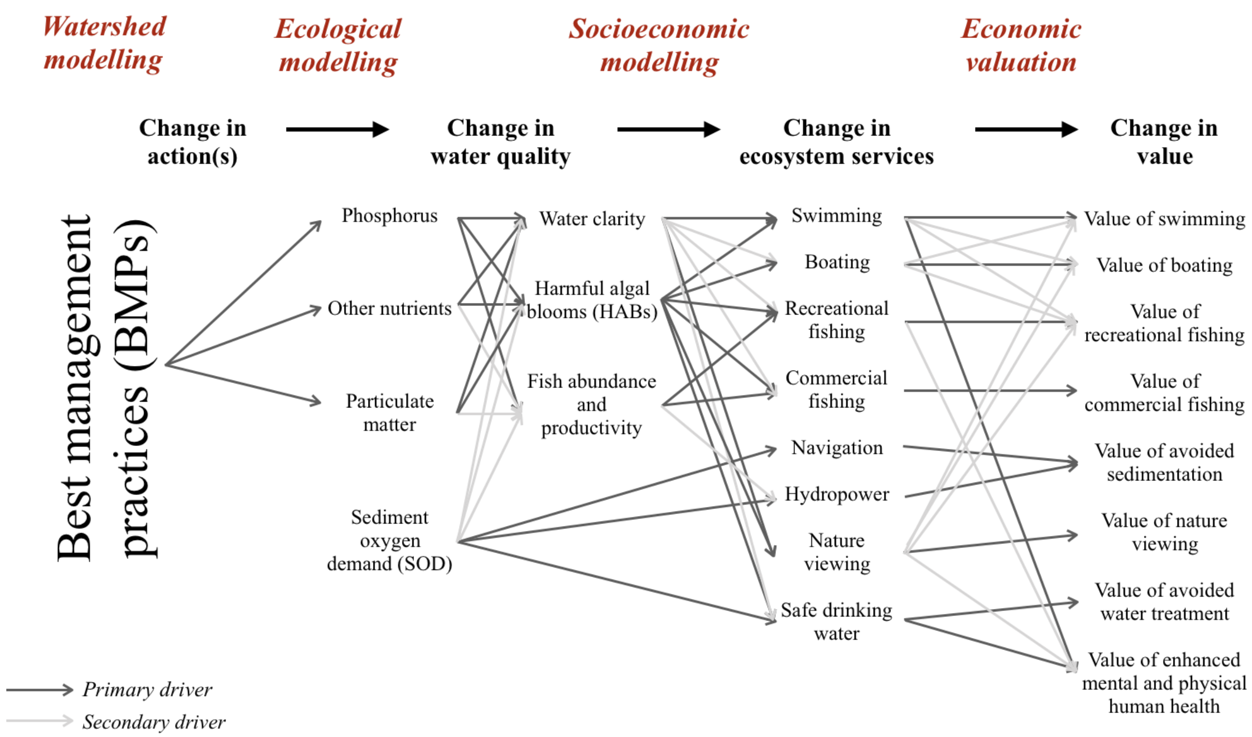


Figure 11: Relationships among human actions, water quality changes, multiple ecosystem goods and services, and associated changes in values in Lake Erie. The proposed addition of a socioeconomic component to the existing integrated watershed-receiving waterbody models will allow the rigorous evaluation of conservation actions and identification of options that allocate financial incentives cost-effectively by funding practices with high predicted environmental benefits per dollar invested.

Ecosystem service valuation can facilitate the active involvement of stakeholders and allow for new insights and knowledge to be passed into the decision-making process. This can be particularly helpful in Lake Erie given its complex ecology and diverse stakeholder groups with divergent goals, priorities, and values (Egoh et al., 2007). Integrating scientific knowledge with ecosystem service values can promote knowledge co-production and co-learning among technical experts, stakeholders, and decision-makers (Laniak et al., 2013). Fortunately, the wealth of watershed and aquatic ecosystem models in Lake Erie offer an excellent foundation upon which relationships among human actions, water quality trends, multiple ecosystem goods and services, and associated changes in values can be depicted (Fig. 11). A characteristic example of the insights that could be gained by such an integrated modelling framework was the study presented by Roy et al. (2010, 2011), which examined the likelihood to find an optimal balance between the conflicting interests of two societal groups, “food producers” and “recreational water users”, in

Sandusky Bay. The latter group includes coastal homeowners, recreational lake users, and local firms that serve recreational lake users, whereas the former represents agricultural operations that generate revenues by activities that increased lake eutrophication. The proposed addition of a socioeconomic component to the existing integrated watershed-receiving waterbody models will allow the rigorous evaluation of conservation actions and identification of options that allocate financial incentives (direct payments, tax credits, insurance, and stewardship certification benefits) cost-effectively by funding practices with high predicted environmental benefits per dollar invested (Palm-Forster et al., 2016).

Consistent with our criticism regarding the skill assessment of the existing modelling work against aggregated spatiotemporal (seasonal/annual, basin- or lake-wide) resolution, we also question the adequacy of the coarse scales selected to establish nutrient loading targets and water quality indicators in Lake Erie (Scavia et al., 2016b,c). This strategy is neither reflective of the range of spatiotemporal dynamics typically experienced in the system nor does it allow to evaluate our progress with ecosystem services at the degree of granularity required to assess the public sentiment. It would seem paradoxical to expect a single-valued standard, based on monitoring and modelling of offshore waters, to capture the water quality conditions in nearshore areas of high public exposure (e.g., beaches). The degree of public satisfaction is primarily determined by the prevailing conditions at a particular recreational site and given date, and not by the average water quality over the entire basin (or lake) and growing season. In our view, the problems with the outdated practice to basing the water quality assessment on the offshore zone with a coarse time scale are twofold: (*i*) we cannot effectively track the progress with the response of the system, as it is not clear to what extent an incremental improvement in the open waters is translated into distinct changes in the nearshore; and (*ii*) the environment targets and decisions are implicitly disconnected with our aspiration to protect ecosystem services and gauge public satisfaction at the appropriate resolution. In the context of adaptive management implementation, we believe that the critical next steps involve the determination of appropriate metrics and scales of expression along with the design of a monitoring program that will allow to effectively track the progress of the system in both time and space (Table S2). Depending on the *ERI* considered, there are different areas for future augmentation in order to more comprehensively monitor the response of Lake Erie. In particular, the assessment of the trophic status may be more appropriate to revolve around extreme (or maximum allowable) phytoplankton or *TP* levels and must explicitly accommodate all the sources of uncertainty by permitting a realistic frequency of violations (Arhonditsis et al., 2016; Kim et al., 2018b). Rather than any type of data averaging, we advocate the assessment of compliance against the proposed probabilistic criteria using daily snapshots collected regularly from different sites during the growing season. The development of the “cyanobacteria index” is certainly useful, but given the technical limitations of the satellite images, we also need other *cHAB* proxy variables that will be collected regularly from the system, including toxins (e.g., Microcystin-LR). The established thresholds for drinking water (1.5 $\mu g\ L^{-1}$) and recreational purposes (20 $\mu g\ L^{-1}$) offer easily defensible targets to track the frequency of compliance of Lake Erie in time and space. Regarding the hypoxia and *Cladophora ERI*s, given our limited mechanistic and quantitative understanding of the primary driving factors, we also propose the development of systematic records for variables that represent direct causal factors of the actual problem, such as phosphorus content in the *Cladophora* tissues, characterization of the organic matter and phosphorus fractionation in the sediments, are the most prudent strategy to move forward.

6 Conclusions

With a wealth of models developed, the next steps of the modelling enterprise should be strategically designed to serve the aspiration of a sustainable resource management in Lake Erie. Rather than "reinventing the wheel" by building new models that bear significant similarity to the ones that are already in place (Mooij et al., 2010), it is critical to craft augmentations that will effectively complement the existing work. In particular, the presence of multiple *SWAT* applications provides assurance that a wide array of physical, chemical, and biological processes with distinct characterizations are considered to reproduce the patterns of flow and nutrient export in agricultural settings, like the Maumee River watershed. While there are models with mechanistically more advanced representation of certain facets of the hydrological cycle (surface runoff, groundwater and sediment erosion) or better equipped to depict urban environments (e.g., *MIKE SHE*, *SWMM*), we believe that greater insights will be gained by revisiting several influential assumptions (tile drainage, fertilizer/manure application rates, *LULC* data) and recalibrating the existing applications to capture both baseline and event-flow conditions and daily nutrient concentration (not loading) variability in multiple locations rather than a single downstream site. Of equal importance is to redesign the land-use management scenarios to accommodate recent conceptual and technical advancements of the life-cycle effectiveness of various *BMP*s, the variability in their starting operational efficiency, and differential response to storm events or seasonality. One of the focal points should also be the role of legacy *P* along with the hydrological and biotransformation mechanisms that modulate *DRP* loading trends. The assessment of the flow-concentration patterns for *N* species and the characterization of processes associated with the nitrogen cycle are still missing in the Lake Erie basin, even though nitrogen could be one of the regulatory factors of the downstream water quality conditions; especially the composition of the algal community. Coupled with the mechanistic tools for the Maumee River watershed, empirical (*SPARROW*-like) models geared to depict basin-specific (rather than continental or regional) nutrient loading conditions can offer a multitude of complementary benefits, such as validate the spatial delineation of hot-spots with higher propensity for nutrient export, narrow down the uncertainty of processes/fluxes parameterized by mechanistic models, and obtain predictive statements constrained within the bounds of data-based parameter estimation.

Counter to the watershed modelling framework for the Maumee River watershed, the multi-model approach for Lake Erie included both data-oriented and process-based models to examine the *ERI* achievability under different nutrient loading conditions. The former models

(*UM-GLERL* and *NOAA* Western Basin *HAB* models) established causal linkages between c*HAB* proxies and external phosphorus loading. Their foundation upon statistical parameter estimation allows for rigorous predictive uncertainty assessment, and thus they represent a pragmatic means to draw forecasts regarding the severity of c*HAB*s. Two critical next steps to further augment the empirical modelling work is the iterative updating as more data are acquired through monitoring and the introduction of other predictors that likely favor the occurrence of cyanobacteria-dominated blooms. After all, while the availability of phosphorus may hierarchically be one of the primary conditions for cyanobacteria dominance, there are several other factors (e.g., nitrogen, iron, light availability, water column stability, and water temperature) that can ultimately determine the winners of the inter-specific competition within the phytoplankton assemblage (Kelly et al., 2018). Because the majority of the process-based models (*ELCOM-CAEDYM, WLEEM, EcoLE*) are far from being constrained by the available data, their primary use has been (and should continue to be) as heuristic tools to advance our understanding of the lake functioning (e.g., potential role of dreissenids, relative importance of meteorological forcing vis-à-vis nutrient availability on the severity of hypoxia), whereas their predictive power is still under question.

With respect to the load-response curves presented by Scavia et al. (2016a,b), the forecasting exercise related to the overall summer phytoplankton biomass in the western basin has a lot of potential to assist the local management efforts. The next augmentations should focus on the development of more reliable empirical model(s) that will connect chlorophyll *a* with a suite of significant predictors, and the advancement of the representation of several factors that could modulate the phytoplankton response to external nutrient loading reductions, such as the degree of reliance of phytoplankton growth upon internal nutrient sources (e.g., microbially mediated regeneration, *P* loading from the sediments), or the strength of top-down control. The coupling of empirical and process-based models to predict the c*HAB* likelihood of occurrence under reduced loading conditions offers a robust foundation to evaluate competing hypotheses and advance our knowledge on the suite of factors that may trigger cyanobacteria dominance in Lake Erie. It is important to recognize though that the reported range of cumulative Maumee March–July annual loads of 1679–2170 MT for achieving the c*HAB* target is likely narrow and does not capture the actual uncertainty with this *ERI*. We also remain skeptical with the optimistic projections of the extent and duration of hypoxia, given our limited knowledge of the sediment diagenesis processes in the central basin and the lack of data related to the vertical profiles of organic matter and phosphorus fractionation or sedimentation/burial rates. Without this piece of information is practically impossible to quantitatively characterize feedback loops of elevated internal loading and sediment oxygen demand, even when the prevailing conditions in the water column are improved, and thus offer strategic foresights into the likelihood to experience a delayed response of the sediments to reduced nutrient loading. It is important to keep in mind that one of the pillars of adaptive management is resilience thinking by monitoring existing problems, highlighting emerging threats, and redefining the research agenda (Johnson et al., 2013; Cook et al., 2014). In terms of the beach fouling by *Cladophora* blooms, the current modelling efforts have been insightful but further enhancement of their predictive value requires a high-resolution study of the northeastern nearshore zone to elucidate the relationships among abiotic conditions, internal *P* content, and sloughing rates in the local mats.

From a management standpoint, it is important to note that the complex mechanistic models are an absolutely worthwhile activity and will continue to assist the on-going management efforts in a meaningfully way. Consistent with Anderson's (2006) views, we believe that prediction is not everything. Even if the structure of complex mathematical models reduces their predictive power or even the ability to conduct rigorous uncertainty analysis, they still offer excellent platforms to gain insights into the direct, indirect, and synergistic effects of the ecological mechanisms forming the foundation of system behavior (Arhonditsis, 2009). For example, the virtual *3D* environment created by *ELCOM-CAEDYM* and/or *WLEEM* can offer a convenient platform to reconcile the coarse-scale (practically offshore) predictions, required to assess the *ERI* achievability, with the granularity that necessitates to elucidate nearshore processes and associated ecosystem services. Even more so, their dynamic integration with the watershed modelling framework will allow to trace the fate of nutrients and suspended solids transported by the Maumee River (and other major tributaries), and generate hypotheses about their impact on the timing and locations where structural shifts in the algal assemblage may occur. Furthermore, being an integral part of the iterative monitoring-modelling-assessment cycles, the foundation of the mechanistic modelling work in Lake Erie can be optimized through reduction of the uncertainty of critical ecological processes or refinement of their structure (e.g., mathematical reformulation of highly sensitive terms, exclusion of irrelevant mechanisms and inclusion of missing ones), thereby augmenting their ability to support ecological forecasts (Arhonditsis et al., 2007). It is thus critical that one of the priorities of the research agenda should be to maintain the ensemble character of the modelling work in Lake Erie. The wide variety of models that have been developed to understand the major causal linkages/ecosystem processes underlying the local water quality problems are a unique feature that should be embraced and further consolidated.

Our analysis questioned the adequacy of the coarse spatiotemporal (seasonal/annual, basin- or lake-wide) scales characterizing both the modelling enterprise and water quality management objectives in Lake Erie. More than anything else, this strategy seems somewhat disconnected from the ecosystem services targeted under Annex 4 of *GLWQA*. In the same context, we argued that ecosystem service valuation can facilitate the decision-making process by identifying cost-effective restoration actions, as we track the evolution of the system over time. While adaptive management and ecosystem service valuation have not typically been used together in decision-making process, they are exceptionally complementary. Both approaches assess ecological systems empirically and are policy-oriented as they describe management implications for stakeholders (Epanchin-Niell et al., 2018). To advance the operationalization of this integrative approach will however require greater interaction among different types of experts of methods, models, and data in social, economic, and environmental sciences. Applying an integrated adaptive management-ecosystem services framework places a premium on articulating policy trade-offs, and therefore has the potential to facilitate the management decisions in the face of uncertainty.

7 References

Alexander, R.B., Smith, R.A., Schwarz, G.E., 2004. Estimates of diffuse phosphorus sources in surface waters of the United States using a spatially referenced watershed model. Water Sci. Technol. 49, 1–10.

Ali, G., Oswald, C.J., Spence, C., Cammeraat, E.L.H., McGuire, K.J., Meixner, T., Reaney, S.M., 2013. Towards a unified threshold-based hydrological theory: Necessary components and recurring challenges. Hydrol. Process. 27(2), 313–318. https://doi.org/10.1002/hyp.9560.

Allan, J.D., Manning, N.F., Smith, S.D.P., Dickinson, C.E., Joseph, C.A., Pearsall, D.R., 2017. Ecosystem services of Lake Erie: Spatial distribution and concordance of multiple services. J. Great Lakes Res. 43, 678–688. https://doi.org/10.1016/j.jglr.2017.06.001.

Allen, C.R., Fontaine, J.J., Pope, K.L., Garmestani, A.S., 2011. Adaptive management for a turbulent future. J. Environ. Manag. 92, 1339–1345. https://doi.org/10.1016/j.jenvman.2010.11.019.

Allinger, L.E., Reavie, E.D., 2013. The ecological history of Lake Erie as recorded by the phytoplankton community. J. Great Lakes Res. 39(3), 365–382. https://doi.org/10.1016/j.jglr.2013.06.014.

Anderson, N.J., Bugmann, H., Dearing, J.A., Gaillard, M.J., 2006. Linking palaeoenvironmental data and models to understand the past and to predict the future. Trends Ecol. Evol. 21(12), 696–704. https://doi.org/10.1016/j.tree.2006.09.005.

Anderson, T.R., 2005. Plankton functional type modelling: Running before we can walk? J. Plankton Res. 27(11), 1078–1081. https://doi.org/10.1093/plankt/fbi076.

Annis, G.M., Pearsall, D.R., Kahl, K.J., Washburn, E.L., May, C.A., Taylor, R.F., Cole, J.B., Ewert, D.N., Game, E.T., Doran, P.J., 2017. Designing coastal conservation to deliver ecosystem and human well-being benefits. PLoS One 12. https://doi.org/10.1371/journal.pone.0172458.

Arhonditsis, G., Kim, D.-K., Kelly, N., Neumann, A., Javed, A., 2018. Uncertainty analysis by Bayesian inference, in: Recknagel, F., Michener, W.K. (Ed.), Ecological Informatics: Data Management and Knowledge Discovery. pp. 215–249. https://doi.org/10.1007/978-3-319-59928-1_11.

Arhonditsis, G.B., Kim, D.K., Shimoda, Y., Zhang, W., Watson, S., Mugalingam, S., Dittrich, M., Geater, K., McClure, C., Keene, B., Morley, A., Richards, A., Long, T., Rao, Y.R., Kalinauskas, R., 2016. Integration of best management practices in the Bay of Quinte watershed with the phosphorus dynamics in the receiving waterbody: What do the models predict? Aquat. Ecosyst. Health 19, 1–18. https://doi.org/10.1080/14634988.2016.1130566.

Arhonditsis, G., Stremilov, S., Gudimov, A., Ramin, M., Zhang, W., 2011. Integration of Bayesian inference techniques with mathematical modeling. In: Wolanski, E., McLusky, D. (Ed.), Treatise on Estuarine and Coastal Science. Elsevier Inc. https://doi.org/10.1016/B978-0-12-374711-2.00910-4.

Arhonditsis, G.B., 2009. Useless arithmetic? Lessons learnt from aquatic biogeochemical modeling. In: Hanrahan, G. (Ed.), Modeling of Pollutants in Complex Environmental Systems. ILM Publications, Hertfordshire, UK (512 pp.).

Arhonditsis, G.B., Papantou, D., Zhang, W., Perhar, G., Massos, E., Shi, M., 2008a. Bayesian calibration of mechanistic aquatic biogeochemical models and benefits for environmental management. J. Marine Syst. 73, 8–30. https://doi.org/10.1016/j.jmarsys.2007.07.004.

Arhonditsis, G.B., Perhar, G., Zhang, W., Massos, E., Shi, M., Das, A., 2008b. Addressing equifinality and uncertainty in eutrophication models. Water Resour. Res. 44, W01420. https://doi.org/10.1029/2007WR005862.

Arhonditsis, G.B., Qian, S.S., Stow, C.A., Lamon, E.C., Reckhow, K.H., 2007. Eutrophication risk assessment using Bayesian calibration of process-based models: Application to a mesotrophic lake. Ecol. Model. 208, 215–229. https://doi.org/10.1016/j.ecolmodel.2007.05.020.

Arhonditsis, G.B., Adams-Vanharn, B.A., Nielsen, L., Stow, C.A., Reckhow, K.H., 2006. Evaluation of the current state of mechanistic aquatic biogeochemical modeling: Citation analysis and future perspectives. Environ. Sci. Technol. 40(21), 6547–6554. https://doi.org/10.1021/es061030q.

Arhonditsis, G.B., Brett, M.T., 2004. Evaluation of the current state of mechanistic aquatic biogeochemical modeling. Mar. Ecol. Prog. Ser. 271, 13–26. https://doi.org/10.1021/es061030q

Arhonditsis, G.B., Winder, M., Brett, M.T., Schindler, D.E., 2004. Patterns and mechanisms of phytoplankton variability in Lake Washington (USA). Water Res. 38, 4013–27. https://doi.org/10.1016/j.watres.2004.06.030.

Arnold, J.G., Srinivasan, R., Muttiah, R.S., Williams, J.R., 1998. Large area hydrologic modeling and assessment. Part I: Model development1. J. Am. Water Resour. Assoc. 34, 73–89. https://doi.org/10.1111/j.1752-1688.1998.tb05961.x.

Arnott, D.L., Vanni, M.J., 1996. Nitrogen and phosphorus recycling by the zebra mussel (Dreissena polymorpha) in the western basin of Lake Erie. Can. J. Fish. Aquat. Sci. 53, 646–659. https://doi.org/10.1139/cjfas-53-3-646.

Auer, M.T., Tomlinson, L.M., Higgins, S.N., Malkin, S.Y., Howell, E.T., Bootsma, H.A., 2010. Great Lakes Cladophora in the 21st century: Same algae-different ecosystem. J. Great Lakes Res. 36, 248–255. https://doi.org/10.1016/j.jglr.2010.03.001.

Auer, M.T., Canale, R.P., 1982. Ecological studies and mathematical modeling of Cladophora in Lake Huron: 3. The dependence of growth rates on internal phosphorus pool size. J. Great Lakes Res. 8, 93–99. https://doi.org/10.1016/S0380-1330(82)71947-1.

Baker, D.B., Johnson, L.T., Confesor, R.B., Crumrine, J.P., 2017. Vertical stratification of soil phosphorus as a concern for dissolved phosphorus runoff in the Lake Erie basin. J. Environ. Qual. 46(6), 1287–1295. https://doi.org/10.2134/jeq2016.09.0337.

Balin, D., Lee, H., Rode, M., 2010. Is point uncertain rainfall likely to have a great impact on distributed complex hydrological modeling? Water Resour. Res. 46, W11520. https://doi.org/10.1029/2009WR007848.

Baver L.D., Gardner, W.H., Gardner, W.R., 1972. Soil Physics. Wiley and Sons, New York, NY, USA.

Beletsky, D., Schwab, D.J., 2001. Modeling circulation and thermal structure in Lake Michigan: Annual cycle and interannual variability. J. Geophys. Res. 106, 19745–19771. https://doi.org/10.1029/2000JC000691.

Belov, A.P., Giles, J.D., 1997. Dynamical model of buoyant cyanobacteria. Hydrobiologia. 349, 87–97. https://doi.org/10.1023/A:1003049629490.

Bertani, I., Obenour, D.R., Steger, C.E., Stow, C.A., Gronewold, A.D., Scavia, D., 2016. Probabilistically assessing the role of nutrient loading in harmful algal bloom formation in western Lake Erie. J. Great Lakes Res. 42, 1184–1192. https://doi.org/10.1016/j.jglr.2016.04.002.

Beven, K. 2006. A manifesto for the equifinality thesis. J. Hydrol. 320, 18–36. https://doi.org/10.1016/j.jhydrol.2005.07.007.

Beven, K., Freer, J., 2001. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. J. Hydrol. 249, 11–29. https://doi.org/10.1016/S0022-1694(01)00421-8.

Bocaniov, S.A., Leon, L.F., Rao, Y.R., Schwab, D.J., Scavia, D., 2016a. Simulating the effect of nutrient reduction on hypoxia in a large lake (Lake Erie, USA-Canada) with a three-dimensional lake model. J. Great Lakes Res. 42, 1228–1240. https://doi.org/10.1016/j.jglr.2016.06.001.

Bocaniov, S.A., Scavia, D., 2016b. Temporal and spatial dynamics of large lake hypoxia: Integrating statistical and three-dimensional dynamic models to enhance lake management criteria. Water Resour. Res. 52, 4247–4263. https://doi.org/10.1002/2015WR018170.

Bocaniov, S.A., Smith, R.E.H., Spillman, C.M., Hipsey, M.R., Leon, L.F., 2014. The nearshore shunt and the decline of the phytoplankton spring bloom in the Laurentian Great Lakes: Insights from a three-dimensional lake model. Hydrobiologia 731, 151–172. https://doi.org/10.1007/s10750-013-1642-2.

Boegman, L., Loewen, M.R., Hamblin, P.F., Culver, D.A., 2001. Application of a two-dimensional hydrodynamic reservoir model to Lake Erie. Can. J. Fish. Aquat. Sci. 58, 858–869. https://doi.org/10.1139/cjfas-58-5-858.

Boegman, L., Loewen, M.R., Culver, D. A., Hamblin, P.F., Charlton, M.N., 2008. Spatial-dynamic modeling of algal biomass in Lake Erie: Relative impacts of dreissenid mussels and nutrient loads. J. Environ. Eng. 134, 456–468. https://doi.org/10.1061/(ASCE)0733-9372(2008)134:6(456).

Borah, D.K., Arnold, J.G., Bera, M., Krug, E.C., Liang, X.-Z., 2007. Storm event and continuous hydrologic modeling for comprehensive and efficient watershed simulations. J. Hydrol. Eng. 12, 605–616. https://doi.org/10.1061/(ASCE)1084-0699(2007)12:6(605).

Borah, D.K., Yagow, G., Saleh, A., Barnes, P.L., Rosenthal, W., Krug, E.C., Hauck, L.M., 2006. Sediment and nutrient modeling for TMDL development and implementation. T. ASABE 49, 967–986. https://doi.org/10.13031/2013.21742.

Borah, D.K., Xia, R., Bera, M., 2001. DWSM – A dynamic watershed simulation model for studying agricultural nonpoint source pollution. 2001 ASAE Annu. Meet. https://doi.org/10.13031/2013.7317.

Borah, D.K., Knapp, H.V., Raman, R.K., Lin, S.D., Soong, T.W.D., 1999. HEC—5Q model applied to a lake system in Illinois for water quality evaluations. Water Int. 24, 240–247. https://doi.org/10.1080/02508069908692167.

Borsuk, M.E., Higdon, D., Stow, C.A., Reckhow, K.H., 2001. A Bayesian hierarchical model to predict benthic oxygen demand from organic matter loading in estuaries and coastal zones. Ecol. Model. 143, 165–181.

Bridgeman, T.B., Schloesser, D.W., Krause, A.E., 2006. Recruitment of Hexagenia mayfly nymphs in western Lake Erie linked to environmental variability. Ecol. Appl. 16, 601–611. https://doi.org/10.1890/1051-0761(2006)016[0601:ROHMNI]2.0.CO;2.

Brox, J.A., Kumar, R.C., Stollery K.R., 1996. Willingness to pay for water quality and supply enhancements in the Grand River watershed. Can. Water Resour. J. 21(3), 285–288. https://doi.org/10.4296/cwrj2103275.

Cai, M., Reavie, E.D., 2018. Pelagic zonation of water quality and phytoplankton in the Great Lakes. Limnology 19, 127–140. https://doi.org/10.1007/s10201-017-0526-y.

Canale, R.P., Redder, T., Swiecki, W., Whelan, G., 2010. Phosphorus budget and remediation plan for Big Platte Lake, Michigan. J. Water Resour. Plan. Manag. 136, 576–586. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000071.

Canale, R.P., Auer, M.T., 1982. Ecological studies and mathematical modeling of Cladophora in Lake Huron: 5. Model development and calibration. J. Great Lakes Res. 8, 112–125. https://doi.org/10.1016/S0380-1330(82)71949-5.

Carroll, R.J., Ruppert, D., Stefanski, L.A., Crainiceanu, C.M., 2006. Measurement Error in Nonlinear Models: A Modern Perspective. Journal of the American Statistical Association. https://doi.org/10.1080/00401706.1997.10485096.

Chesapeake Bay Program (CBP). 2015. Protocol for the development, review, and approval of loading and effectiveness estimates for nutrient and sediment controls in the Chesapeake Bay watershed model. Available at: https://www.chesapeakebay.net/documents/CBP_BMP_Expert_Panel_Protocol_WQGIT_approved_7.13.15.pdf.

Chaffin, J.D., Bridgeman, T.B., 2014. Organic and inorganic nitrogen utilization by nitrogen-stressed cyanobacteria during bloom conditions. J. Appl. Phycol. 26, 299–309. https://doi.org/10.1007/s10811-013-0118-0.

Chapra, S.C., Dolan, D.M., Dove, A., 2016. Mass-balance modeling framework for simulating and managing long-term water quality for the lower Great Lakes. J. Great Lakes Res. 42, 1166–1173. https://doi.org/10.1016/j.jglr.2016.04.008.

Chapra, S.C., Sonzogni, W.C., 1979. Great Lakes total phosphorus budget for the mid 1970s. J. Water Pollut. Con. F. 51, 2524–2533.

Chapra, S. 1977a. Total phosphorus model for the Great Lakes. J. Environ. Eng. 103, 147–161.

Chapra, S.C., Robertson, A., 1977b. Great Lakes eutrophication: The effect of point source control of total phosphorus. Science. 196, 1448–1450. https://doi.org/10.1126/science.196.4297.1448.

Chen, M. 2013. Valuation of Public Great Lakes Beaches in Michigan. Ph.D. dissertation, Michigan State University, East Lansing. Available at: http://web2.msue.msu.edu/afreTheses/fulltext/Min%20Chen%20Dissertation.pdf (accessed November 22, 2017).

Cheng, V., Arhonditsis, G.B., Brett, M.T., 2010. A revaluation of lake-phosphorus loading models using a Bayesian hierarchical framework. Ecol. Res. 25, 59–76. https://doi.org/10.1007/s11284-009-0630-5.

Christakos, G., 2003. Critical conceptualism in environmental modeling and prediction. Environ.

Sci. Technol. 37, 4685–4693. https://doi.org/10.1021/es020932y.

Cole, T.M., Wells, S.A., 2006. CE-QUAL-W2: A Two Dimensional, Laterally Averaged, Hydrodynamic and Water Quality Model, Version 3.5. U.S. Army Corps of Engineers, Washington, DC.

Conroy, J.D., Boegman, L., Zhang, H., Edwards, W.J., Culver, D.A., 2011. "Dead Zone" dynamics in Lake Erie: The importance of weather and sampling intensity for calculated hypolimnetic oxygen depletion rates. Aquat. Sci. 73, 289–304. https://doi.org/10.1007/s00027-010-0176-1.

Conroy, J.D., Edwards, W.J., Pontius, R.A., Kane, D.D., Zhang, H., Shea, J.F., Richey, J.N., Culver, D.A., 2005. Soluble nitrogen and phosphorus excretion of exotic freshwater mussels (Dreissena spp.): potential impacts for nutrient remineralisation in western Lake Erie. Freshw. Biol. 50(7), 1146–1162.

Cook, C.N., Inayatullah, S., Burgman, M.A., Sutherland, W.J., Wintle, B.A., 2014. Strategic foresight: How planning for the unpredictable can improve environmental decision-making. Trends Ecol. Evol. 29, 531–541. https://doi.org/10.1016/j.tree.2014.07.005.

Costanza, R., D'Arge, R., De Groot, R., Farber, S., Grasso, M., Hannon, B., Limburg, K., Naeem, S., O'Neill, R. V., Paruelo, J., Raskin, R.G., Sutton, P., Van Den Belt, M., 1997. The value of the world's ecosystem services and natural capital. Nature 387, 253–260. https://doi.org/10.1038/387253a0.

Cousino, L.K., Becker, R.H., Zmijewski, K.A., 2015. Modeling the effects of climate change on water, sediment, and nutrient yields from the Maumee River watershed. J. Hydrol. Reg. Stud. 4, 762–775. https://doi.org/10.1016/j.ejrh.2015.06.017.

Crossman, J., Futter, M.N., Oni, S.K., Whitehead, P.G., Jin, L., Butterfield, D., Baulch, H.M., Dillon, P.J., 2013. Impacts of climate change on hydrology and water quality: Future proofing management strategies in the Lake Simcoe watershed, Canada. J. Great Lakes Res. 39, 19–32. https://doi.org/10.1016/j.jglr.2012.11.003.

Culbertson, A.M., Martin, J.F., Aloysius, N., Ludsin, S.A., 2016. Anticipated impacts of climate change on 21st century Maumee River discharge and nutrient loads. J. Great Lakes Res. 42, 1332–1342. https://doi.org/10.1016/j.jglr.2016.08.008.

Daloğlu, I., Cho, K.H., Scavia, D., 2012. Evaluating causes of trends in long-term dissolved reactive phosphorus loads to Lake Erie. Environ. Sci. Technol. 46, 10660–10666. https://doi.org/10.1021/es302315d.

Davis, T.W., Bullerjahn, G.S., Tuttle, T., McKay, R.M., Watson, S.B., 2015. Effects of increasing nitrogen and phosphorus concentrations on phytoplankton community growth and toxicity during Planktothrix blooms in Sandusky Bay, Lake Erie. Environ. Sci. Technol. 49, 7197–7207. https://doi.org/10.1021/acs.est.5b00799.

Davis T.W. Koch F. Marcoval M.A. Wilhelm S.W. Gobler C.J., 2012. Mesozooplankton and microzooplankton grazing during cyanobacterial blooms in the western basin of Lake Erie. Harmful Algae 15, 26–35. https://doi:10.1016/j.hal.2011.11.002

Depew, D.C., Houben, A.J., Guildford, S.J., Hecky, R.E., 2011. Distribution of nuisance Cladophora in the lower Great Lakes: Patterns with land use, near shore water quality and dreissenid abundance. J. Great Lakes Res. 37, 656–671. https://doi.org/10.1016/j.jglr.2011.08.011.

Dietz, M.E., 2007. Low impact development practices: A review of current research and recommendations for future directions. Water. Air. Soil Pollut. 186, 351–363. https://doi.org/10.1007/s11270-007-9484-z.

Dillaha, T.A., Reneau, R.B., Mostaghimi, S., Lee, D., 1989. Vegetative filter strips for agricultural nonpoint source pollution control. Trans. Am. Soc. Agric. Eng. 32, 513–519. https://doi.org/10.13031/2013.31033.

Di Toro, D.M., Thomas, N.A., Herdendorf, C.E., Winfield, R.P., Connolly, J.P., 1987. A post audit of a Lake Erie eutrophication model. J. Great Lakes Res. 13, 801–825. https://doi.org/10.1016/S0380-1330(87)71692-X.

Dittrich, M., Chesnyuk, A., Gudimov, A., McCulloch, J., Quazi, S., Young, J., Winter, J., Stainsby, E., Arhonditsis, G., 2013. Phosphorus retention in a mesotrophic lake under transient loading conditions: Insights from a sediment phosphorus binding form study. Water Res. 47, 1433–1447. https://doi.org/10.1016/j.watres.2012.12.006.

Doan, P.T.K., Watson, S.B., Markovic, S., Liang, A., Guo, J., Mugalingam, S., Stokes, J., Morley, A., Zhang, W., Arhonditsis, G.B., Dittrich, M., 2018. Phosphorus retention and internal loading in the Bay of Quinte, Lake Ontario, using diagenetic modelling. Sci. Total Environ. 636, 39–51. https://doi.org/10.1016/j.scitotenv.2018.04.252.

Dodd, R.J., Mcdowell, R.W., Condron, L.M., 2013. Changes in soil phosphorus availability and potential phosphorus loss following cessation of phosphorus fertiliser inputs. Soil Res. 51, 427–436. https://doi.org/10.1071/SR13168.

Dong, F., Liu, Y., Wu, Z., Chen, Y., Guo, H., 2018. Identification of watershed priority management areas under water quality constraints: A simulation-optimization approach with ideal load reduction. J. Hydrol. 562, 577–588. https://doi.org/10.1016/j.jhydrol.2018.05.033.

Dorfman, J.H., Keeler, A.G., Kriesel, W., 1996. Valuing risk-reducing interventions with hedonic models: The case of erosion protection. J. Agric. Resour. Econ. 21, 109–119.

Dorioz, J.M., Wang, D., Poulenard, J., Trévisan, D., 2006. The effect of grass buffer strips on phosphorus dynamics-A critical review and synthesis as a basis for application in agricultural landscapes in France. Agric. Ecosyst. Environ. 117, 4–21. https://doi.org/10.1016/j.agee.2006.03.029.

Dröscher, I., Patoine, A., Finlay, K., Leavitt, P.R., 2009. Climate control of the spring clear-water phase through the transfer of energy and mass to lakes. Limnol. Oceanogr. 54, 2469–2480. https://doi.org/10.4319/lo.2009.54.6_part_2.2469.

Dove, A., Chapra, S.C., 2015. Long-term trends of nutrients and trophic response variables for the Great Lakes. Limnol. Oceanogr. 60, 696–721. https://doi.org/10.1002/lno.10055.

Edwards, P.J., Wood, F., Quinlivan, R.L., 2016. Effectiveness of best management practices that have application to forest roads: a literature synthesis. USDA FS Northern Research Station (1719 pp.). Available at: https://www.fs.fed.us/nrs/pubs/gtr/gtr_nrs163.pdf.

Egoh, B., Rouget, M., Reyers, B., Knight, A.T., Cowling, R.M., van Jaarsveld, A.S., Welz, A., 2007. Integrating ecosystem services into conservation assessments: A review. Ecol. Econ. 63, 714–721. https://doi.org/10.1016/j.ecolecon.2007.04.007.

Epanchin-Niell, R.S., Boyd, J.W., Macauley, M.K., Scarlett, L., Shapiro, C.D., Williams, B.K. 2018. Integrating adaptive management and ecosystem services concepts to improve natural resource management— Challenges and opportunities. U.S. Geological Survey Circular 1439.

Reston, Virginia. https://doi.org/10.3133/cir1439.

Fang, X., Stefan, H.G., 2009. Simulations of climate effects on water temperature, dissolved oxygen, and ice and snow covers in lakes of the contiguous U.S. under past and future climate scenarios. Limnol. Oceanogr. 54, 2359–2370. https://doi.org/10.4319/lo.2009.54.6_part_2.2359.

Farber, S.C., Costanza, R., Wilson, M.A., 2002. Economic and ecological concepts for valuing ecosystem services. Ecol. Econ. 41, 375–392. https://doi.org/10.1016/S0921-8009(02)00088-5.

Fennel, W., Neumann, T., 2003. Variability of copepods as seen in a coupled physical–biological model of the Baltic Sea. ICES Mar. Sci. Symp. 219, 208–219.

Finlay, K., Patoine, A., Donald, D.B., Bogard, M.J., Leavitt, P.R., 2010. Experimental evidence that pollution with urea can degrade water quality in phosphorus-rich lakes of the Northern Great Plains. Limnol. Oceanogr. 55, 1213–1230. https://doi.org/10.4319/lo.2010.55.3.1213.

Gebhardt, M.R., Daniel, T.C., Schweizer, E.E., Allmaras, R.R., 1985. Conservation Tillage. Science (80). 230, 625–630. https://doi.org/10.1126/science.230.4726.625.

Gildow, M., Aloysius, N., Gebremariam, S., Martin, J., 2016. Fertilizer placement and application timing as strategies to reduce phosphorus loading to Lake Erie. J. Great Lakes Res. 42, 1281–1288. https://doi.org/10.1016/j.jglr.2016.07.002.

Gilpin, A., 2004. Environmental economics: A critical overview. Wiley Chichester, UK.

Great Lakes Water Quality Agreement (GLWQA), 2016. The United States and Canada adopt phosphorus load reduction targets to combat Lake Erie algal blooms. Available at: https://binational.net/2016/02/22/finalptargets-ciblesfinalesdep/.

Godsey, S.E., Kirchner, J.W., Clow, D.W., 2009. Concentration–discharge relationships reflect chemostatic characteristics of US catchments. Hydrol. Process. 23, 1844–1864. http://dx.doi.org/10.1002/hyp.7315.

Goldman, J.C., 1984. Oceanic nutrient cycles. In: Fasham, M.J. (Ed.), Flows of energy and materials in marine ecosystems: Theory and practice. Plenum, London, UK (743 pp.).

Goodess, C.M., 2013. How is the frequency, location and severity of extreme events likely to change up to 2060? Environ. Sci. Policy 27, S4–S14. https://doi.org/10.1016/j.envsci.2012.04.001.

Green, T.R., Salas, J.D., Martinez, A., Erskine, R.H., 2007. Relating crop yield to topographic attributes using spatial analysis neural networks and regression. Geoderma 139, 23–37. https://doi.org/10.1016/j.geoderma.2006.12.004.

Griffin, C.B., 1995. Uncertainty analysis of BMP effectiveness for controlling nitrogen from urban nonpoint sources. J. Am. Water Resour. Assoc. 31, 1041–1050. https://doi.org/10.1111/j.1752-1688.1995.tb03420.x.

Gulati, R.D., Perez Martinez, C., Siewertsen, K., 1995. Zooplankton as a compound mineralising and synthesizing system: phosphorus excretion. Hydrobiologia 315, 25–37. https://doi.org/10.1007/BF00028628.

Gudimov, A., O'Connor, E., Dittrich, M., Jarjanazi, H., Palmer, M.E., Stainsby, E., Winter, J.G., Young, J.D., Arhonditsis, G.B., 2012. Continuous Bayesian network for studying the causal links between phosphorus loading and plankton patterns in Lake Simcoe, Ontario, Canada.

Environ. Sci. Technol. 46, 7283–7292. https://doi.org/10.1021/es300983r.

Gudimov, A., Kim, D.K., Young, J.D., Palmer, M.E., Dittrich, M., Winter, J.G., Stainsby, E., Arhonditsis, G.B., 2015. Examination of the role of dreissenids and macrophytes in the phosphorus dynamics of Lake Simcoe, Ontario, Canada. Ecol. Inform. 26, 36–53. https://doi.org/10.1016/j.ecoinf.2014.11.007.

Gudimov, A., McCulloch, J., Chen, J., Doan, P., Arhonditsis, G., Dittrich, M., 2016. Modeling the interplay between deepwater oxygen dynamics and sediment diagenesis in a hard-water mesotrophic lake. Ecol. Inform. 31, 59–69. https://doi.org/10.1016/j.ecoinf.2015.11.005.

Hamrick, J.M., 1992. Estuarine environmental impact assessment using a three-dimensional circulation and transport model, in: Proceedings of the 2nd International Conference on Estuarine Coastal Model. pp. 292–303.

Harke, M.J., Steffen, M.M., Gobler, C.J., Otten, T.G., Wilhelm, S.W., Wood, S.A., Paerl, H.W., 2016. A review of the global ecology, genomics, and biogeography of the toxic cyanobacterium, Microcystis spp. Harmful Algae. 54, 4–20. https://doi.org/10.1016/j.hal.2015.12.007.

Harris T.D., V.H. Smith, 2016. Do persistent organic pollutants stimulate cyanobacterial blooms? Inland Waters 6, 124-130, https://doi.org/10.5268/IW-6.2.887

Hayder, S., 2014. Socio-economic impact of the presence of Asian Carp in the Great Lakes basin. Policy and Economics, Fisheries and Oceans Canada, Winnipeg, MB. iv C 76 pp. Available at: http://publications.gc.ca/collections/collection_2015/mpo-dfo/Fs23-600-2014-eng.pdf.

He, C. and C. DeMarchi. 2010. Modeling spatial distributions of point and nonpoint source pollution loadings in the Great Lakes watersheds. Int. J. Environ. Sci. Eng. 2(1), 24–30.

Higgins, S.N., Howell, E.T., Hecky, R.E., Guildford, S.J., Smith, R.E., 2005. The wall of green: The status of Cladophora glomerata on the northern shores of Lake Erie's eastern basin, 1995-2002. J. Great Lakes Res. 31, 547–563. https://doi.org/10.1016/S0380-1330(05)70283-5.

Higgins, S.N., Hecky, R.E., Guildford, S.J., 2006. Environmental controls of Cladophora growth dynamics in eastern Lake Erie: Application of the Cladophora growth model (CGM). J. Great Lakes Res. 32, 629–644. https://doi.org/10.3394/0380-1330(2006)32.

Higgins, S.N., Malkin, S.Y., Todd Howell, E., Guildford, S.J., Campbell, L., Hiriart-Baer, V., Hecky, R.E., 2008. An ecological review of Cladophora glomerata (Chlorophyta) in the Laurentian Great Lakes. J. Phycol. 44(4), 839–854. https://doi.org/10.1111/j.1529-8817.2008.00538.x.

Hipsey, M.R., 2008a. The CWR Computational Aquatic Ecosystem Dynamics Model CAEDYM. User Manual. Centre for Water Research (CWR), University of Western Australia.

Hipsey, M.R. and Hamilton, D.P., 2008b. Computational aquatic ecosystem dynamic model: CAEDYM ver. 3, Science Manual, Centre for Water Research Report, University of Western Australia.

Ho, J.C., Michalak, A.M., 2015. Challenges in tracking harmful algal blooms: A synthesis of evidence from Lake Erie. J. Great Lakes Res. 41, 317–325. https://doi.org/10.1016/j.jglr.2015.01.001.

Horst, G.P., Sarnelle, O., White, J.D., Hamilton, S.K., Kaul, R.R.B., Bressie, J.D., 2014. Nitrogen availability increases the toxin quota of a harmful cyanobacterium, Microcystis aeruginosa.

Water Res. 54, 188–198. https://doi.org/10.1016/j.watres.2014.01.063.

Huntington, T.G., 2006. Evidence for intensification of the global water cycle: Review and synthesis. J. Hydrol. 319, 83–95. https://doi.org/10.1016/j.jhydrol.2005.07.003.

International Joint Commission, 2014. A balanced diet of Lake Erie: reducing phosphorus loadings and harmful algal blooms. Report of the Lake Erie Ecosystem Priority. ISBN: 978-1-927336-07-6, pp. 96.

Isely, P., Isely, E.S., Hause, C., Steinman, A.D., 2018. A socioeconomic analysis of habitat restoration in the Muskegon Lake area of concern. J. Great Lakes Res. 44, 330–339. https://doi.org/10.1016/j.jglr.2017.12.002.

Jackson, M.B., Hamdy, Y.S., 1982. Projected Cladophora growth in southern Georgian Bay in response to proposed municipal sewage treatment plant discharges to the Mary Ward Shoals. J. Great Lakes Res. 8, 153–163. https://doi.org/10.1016/S0380-1330(82)71953-7.

James, W.P., Warinner, J., Reedy, M., 1992. Application of the Green-Ampt infiltration equation to watershed modeling. J. Am. Water Resour. Assoc. 28, 623–635. https://doi.org/10.1111/j.1752-1688.1992.tb03182.x.

Jarvie, H.P., Johnson, L.T., Sharpley, A.N., Smith, D.R., Baker, D.B., Bruulsema, T.W. and Confesor, R., 2017. Increased soluble phosphorus loads to Lake Erie: Unintended consequences of conservation practices? J. Environ. Qual. 46, 123–132. https://doi.org/10.2134/jeq2016.07.0248.

Jarvie, H.P., Sharpley, A.N., Spears, B., Buda, A.R., May, L., Kleinman, P.J.A., 2013. Water quality remediation faces unprecedented challenges from "legacy phosphorus". Environ. Sci. Technol. 47, 8997–8998. https://doi.org/10.1021/es403160a.

Jarvis, P., Dow, J., Dermott, R., and Bonnell, R. 2000. Zebra (Dreissena polymorpha) and quagga mussel (Dreissena bugensis) distribution and density in Lake Erie, 1992–1998. Canadian Technical Report of Fisheries and Aquatic Sciences No. 2304: i–v; 1–46.

Jiang, L.P., Fang, X., Stefan, H.G., Jacobson, P.C., Pereira, D.L., 2012. Oxythermal habitat parameters and identifying cisco refuge lakes in Minnesota under future climate scenarios using variable benchmark periods. Ecol. Model. 232, 14–27.

Johengen, T.H., Vanderploeg, H.A., Liebig, J.R., 2013. Effects of algal composition, seston composition, and feeding rate on zebra mussel (Dreissena polymorpha) nutrient excretion in two Laurentian Great Lakes. In: Nalepa, T.F., Schloesser, D.W. (Ed.), Quagga and Zebra Mussels: Biology, Impacts, and Control, Second Edition. CRC Press, Boca Raton, Florida, USA (778 pp.).

Johnson, D.D.P., Blumstein, D.T., Fowler, J.H., Haselton, M.G., 2013. The evolution of error: Error management, cognitive constraints, and adaptive decision-making biases. Trends Ecol. Evol. 28, 478–481. https://doi.org/10.1016/j.tree.2013.05.014.

Johnson, F.A., Boomer, G.S., Williams, B.K., Nichols, J.D., Case, D.J., 2015. Multilevel learning in the adaptive management of waterfowl harvests: 20 years and counting. Wildlife Soc. B. 39, 9–19. https://doi.org/10.1002/wsb.518.

Jones, C., Lick, W., 2001. Sediment erosion rates: their movement and use in modeling. Texas A&M Dredging Seminar, College Station TX (15 pp.).

Kalcic, M.M.C., Kirchhoff, C., Bosch, N., Muenich, R.L., Murray, M., Griffith Gardner, J., Scavia, D., 2016. Engaging stakeholders to define feasible and desirable agricultural conservation in western Lake Erie watersheds. Environ. Sci. Technol. 50, 8135–8145. https://doi.org/10.1021/acs.est.6b01420.

Kamarainen, A.M., Penczykowski, R.M., Van de Bogert, M.C., Hanson, P.C. and Carpenter, S.R., 2009. Phosphorus sources and demand during summer in a eutrophic lake. Aquat. Sci. 71, 214–227. https://doi.org/ 10.1007/s00027-009-9165-7.

Kane, D.D., Ludsin, S.A., Briland, R.D., Culver, D.A. and Munawar, M., 2015. Ten+ years gone: Continued degradation of offshore planktonic communities in US waters of Lake Erie's western and central basins (2003–2013). J. Great Lakes Res. 41, 930–933. https://doi.org/ 10.1016/j.jglr.2015.06.002.

Karatayev, A.Y., Burlakova, L.E., Mehler, K., Bocaniov, S.A., Collingsworth, P.D., Warren, G., Kraus, R.T. and Hinchey, E.K., 2017. Biomonitoring using invasive species in a large Lake: Dreissena distribution maps hypoxic zones. J. Great Lakes Res. 44, 639–649. https://doi.org/ 10.1016/j.jglr.2017.08.001.

Karatayev, A.Y., Burlakova, L.E., Pennuto, C., Ciborowski, J., Karatayev, V.A., Juette, P. and Clapsadl, M., 2014. Twenty five years of changes in Dreissena spp. populations in Lake Erie. J. Great Lakes Res. 40, 550–559. https://doi.org/ 10.1016/j.jglr.2014.04.010.

Katsev, S., 2017. When large lakes respond fast: A parsimonious model for phosphorus dynamics. J. Great Lakes Res. 43, 199–204. https://doi.org/ 10.1016/j.jglr.2016.10.012.

Keitzer, S.C., Ludsin, S.A., Sowa, S.P., Annis, G., Arnold, J.G., Daggupati, P., Froehlich, A.M., Herbert, M.E., Johnson, M.V.V., Sasson, A.M. and Yen, H., 2016. Thinking outside of the lake: Can controls on nutrient inputs into Lake Erie benefit stream conservation in its watershed? J. Great Lakes Res. 42, 1322–1331. https://doi.org/ 10.1016/j.jglr.2016.05.012.

Kelch, D., Lichtkoppler, F., Sohngen, B. and Daigneault, A., 2006. The value of steelhead (Onchorhynchus mykiss) angling in Lake Erie tributaries. J. Great Lakes Res. 32, 424–433. https://doi.org/ 10.3394/0380-1330(2006)32[424:TVOSOM]2.0.CO;2.

Kelly, N.E., Javed, A., Shimoda, Y., Zastepa, A., Watson, S., Arhonditsis, G.B., 2018. A Bayesian modelling approach for predicting risks of microcystin toxin concentrations in freshwaters. Environ. Sci. Technol. In submission.

Kieta, K.A., Owens, P.N., Lobb, D.A., Vanrobaeys, J.A. and Flaten, D.N., 2018. Phosphorus dynamics in vegetated buffer strips in cold climates: A review. Environ. Rev. https://doi.org/ 10.1139/er-2017-0077.

Kim, D.-K., Dong, F., Neumann, A., Javed, A., Shimoda, Y., Arhonditsis, G.B., 2018a. Evaluation of the capacity of a swat-based ensemble modelling strategy to guide best management practices. Ecol. Inform. In submission.

Kim, D.K., Ramin, M., Cheng, V., Javed, A., Kaluskar, S., Kelly, N., Kobiliris, D., Neumann, A., Ni, F., Peller, T., Perhar, G., Shimoda, Y., Visha, A., Wellen, C., Yang, C., Mugalingam, S., Arhonditsis, G.B., 2018b. An integrative methodological framework for setting environmental criteria: Evaluation of stakeholder perceptions. Ecol. Inform. In press.

Kim, D.-K, Kaluskar, S., Mugalingam, S., Blukacz-Richards, A., Long, T., Morley, A. and Arhonditsis, G.B., 2017. A Bayesian approach for estimating phosphorus export and delivery rates with the SPAtially Referenced Regression On Watershed attributes (*SPARROW*) model. Ecol. Inform. 37, 77–91. https://doi.org/ 10.1016/j.ecoinf.2016.12.003.

Kim, D.-K., Kaluskar, S., Mugalingam, S., and Arhonditsis, G.B., 2016. Evaluating the relationships between watershed physiography, land use patterns, and phosphorus loading in the Bay of Quinte Basin, Ontario, Canada. J Great Lakes Res. 42, 972-984. http://dx.doi.org/10.1016/j.jglr.2016.07.008

Kim, D.-K, Zhang, W., Watson, S. and Arhonditsis, G.B., 2014. A commentary on the modelling of the causal linkages among nutrient loading, harmful algal blooms, and hypoxia patterns in Lake Erie. J. Great Lakes Res. 40, 117–129. https://doi.org/ 10.1016/j.jglr.2014.02.014.

Kim, D.-K, Zhang, W., Rao, Y.R., Watson, S., Mugalingam, S., Labencki, T., Dittrich, M., Morley, A. and Arhonditsis, G.B., 2013. Improving the representation of internal nutrient recycling with phosphorus mass balance models: a case study in the Bay of Quinte, Ontario, Canada. Ecol. Model. 256, 53–68. https://doi.org/ 10.1016/j.ecolmodel.2013.02.017.

King, K.W., Williams, M.R., Johnson, L.T., Smith, D.R., LaBarge, G.A. and Fausey, N.R., 2017. Phosphorus availability in western Lake Erie Basin drainage waters: Legacy evidence across spatial scales. J. Environ. Qual. 46, 466–469. https://doi.org/ 10.2134/jeq2016.11.0434.

King, K.W., Arnold, J.G. and Bingner, R.L., 1999. Comparison of Green-Ampt and curve number methods on Goodwin Creek watershed using SWAT. T. ASAE. 42, 919–926 https://doi.org/ 10.13031/2013.13272.

Kirnak, H., 2002. Comparison of erosion and runoff predicted by WEPP and AGNPS models using a geographic information system. Turk. J. Agric. For. 26, 261–268.

Klaus, J. and McDonnell, J.J., 2013. Hydrograph separation using stable isotopes: Review and evaluation. J. Hydrol. 505, 47-64. https://doi.org/ 10.1016/j.jhydrol.2013.09.006.

Kleinman, P.J., Sharpley, A.N., McDowell, R.W., Flaten, D.N., Buda, A.R., Tao, L., Bergstrom, L. and Zhu, Q., 2011. Managing agricultural phosphorus for water quality protection: principles for progress. Plant Soil. 349, 169–182. https://doi.org/ 10.1007/s11104-011-0832-9.

Knapp, A.K., Beier, C., Briske, D.D., Classen, A.T., Luo, Y., Reichstein, M., Smith, M.D., Smith, S.D., Bell, J.E., Fay, P.A. and Heisler, J.L., 2008. Consequences of more extreme precipitation regimes for terrestrial ecosystems. BioScience, 58, 811–821. https://doi.org/ 10.1641/B580908.

Kovacic, D.A., David, M.B., Gentry, L.E., Starks, K.M. and Cooke, R.A., 2000. Effectiveness of constructed wetlands in reducing nitrogen and phosphorus export from agricultural tile drainage. J. Environ. Qual. 29, 1262–1274. https://doi.org/ 10.2134/jeq2000.00472425002900040033x.

Kowalezewska-Madura, K., Goldyn, R., Szyper, H., 2007. Zooplankton phosphorus excretion in Swarzedzkie Lake (Western Poland) and its influence on phytoplankton. Oceanol. Hydrobiol. St. 36, 3–16. https://doi.org/ 10.2478/v10009-007-0010-9.

Kriesel, W., Randall, A., Lichtkoppler, F., 1993. Estimating the benefits of shore erosion protection in Ohio's Lake Erie housing market. Water Resour. Res. 29, 795–801. https://doi.org/10.1029/92WR02539.

Lake Erie Improvement Association (LEIA), 2012. Sustaining Healthy Waters for Lake Erie's Economy. Lake Erie Improvement Association Strategic Plan for Lake Erie Partners. Unpublished report (57 pp). Available at: https://studylib.net/doc/7766299/the-strategic-plan---lake-erie-improvement-association.

Lam, D.C., Schertzer, W.M., Fraser, A.S., 1983. Simulation of Lake Erie Water Quality Responses to Loading and Weather Variations. National Water Research Institute, Inland Waters Directorate, Canada Center for Inland Waters, Burlington, Ontario (Scientific series No. 134).

Lam, D.C.L., Schertzer, W.M., McCrimmon, Charlton, M., Millard, S., 2008. Modeling phosphorus and dissolved oxygen conditions pre- and post-Dreissena arrival in Lake Erie. Checking the Pulse of Lake Erie. Backhuys Publ., Leiden, The Netherlands, pp. 97–121.

Laniak, G.F., Olchin, G., Goodall, J., Voinov, A., Hill, M., Glynn, P., Whelan, G., Geller, G., Quinn, N., Blind, M. and Peckham, S., 2013. Integrated environmental modeling: a vision and roadmap for the future. Environ. Modell. Softw. 39, 3–23. https://doi.org/ 10.1016/j.envsoft.2012.09.006.

Leitão, J., et al., 2018. Identifying the best locations to install flow control devices in sewer networks to enable in-sewer storage. J. Hydrol. 556, 371–383. https://doi.org/ 10.1016/j.jhydrol.2017.11.020.

León, L.F., Smith, R.E., Malkin, S.Y., Depew, D., Hipsey, M.R., Antenucci, J.P., Higgins, S.N., Hecky, R.E. and Rao, R.Y., 2012. Nested 3D modeling of the spatial dynamics of nutrients and phytoplankton in a Lake Ontario nearshore zone. J. Great Lakes Res. 38,171-183. https://doi.org/ 10.1016/j.jglr.2012.02.006.

León, L.F., Smith, R.E., Hipsey, M.R., Bocaniov, S.A., Higgins, S.N., Hecky, R.E., Antenucci, J.P., Imberger, J.A. and Guildford, S.J., 2011. Application of a 3D hydrodynamic–biological model for seasonal and spatial dynamics of water quality and phytoplankton in Lake Erie. J. Great Lakes Res. 37, 41–53. https://doi.org/ 10.1016/j.jglr.2010.12.007.

León, L.F., Smith, R.E.H., Romero, J.R., Hecky, R.E., 2006. Lake Erie hypoxia simulations with ELCOM-CAEDYM. In: Proceedings of the 3rd Biennial meeting of the International Environmental Modelling and Software Society, Burlington, Vermont, July 9–13.

León, L.F., Imberger, J., Smith, R.E.H., Hecky, R.E., Lam, D.C.L., Schertzer, W.M., 2005. Modeling as a tool for nutrient management in Lake Erie: A hydrodynamics study. J. Great Lakes Res. 31, Supplement 2, 309–318. https://doi.org/10.1016/S0380-1330(05)70323-3.

Lesht, B.M., Fontaine III, T.D. and Dolan, D.M., 1991. Great Lakes total phosphorus model: post audit and regionalized sensitivity analysis. J. Great Lakes Res. 17, 3–17. https://doi.org/ 10.1016/S0380-1330(91)71337-3.

Li, Y. and Babcock, R.W., 2014. Green roof hydrologic performance and modeling: a review. Water Sci. Technol. 69, 727–738. https://doi.org/ 10.2166/wst.2013.770.

LimnoTech, 2013. Development of an Integrated Modeling Approach for Quantifying the GLRI Deposition Metric: Pilot Application to Toledo Harbor. Technical Report Prepared for USACE-Buffalo District, Buffalo, NY (125 pp.).

LimnoTech, 2010. Development, Calibration, and Application of the Lower Maumee River — Maumee Bay Model. Technical Report Prepared for USACE-Buffalo District,Buffalo, NY (127 pp.).

LimnoTech, 2009. Upper Mississippi River-Lake Pepin water quality model development, calibration, and application. Technical Report Prepared for Minnesota Pollution Control Agency. July 2009. Available at: https://www.pca.state.mn.us/sites/default/files/wq-iw9-01jc.pdf.

Liu, Y., Engel, B.A., Flanagan, D.C., Gitau, M.W., McMillan, S.K., Chaubey, I., Singh, S., 2018.

Modeling framework for representing long-term effectiveness of best management practices in addressing hydrology and water quality problems: Framework development and demonstration using a Bayesian method. J. Hydrol. 560, 530–545. https://doi.org/10.1016/j.jhydrol.2018.03.053.

Liu, Y., Engel, B.A., Flanagan, D.C., Gitau, M.W., McMillan, S.K., Chaubey, I., 2017. A review on effectiveness of best management practices in improving hydrology and water quality: Needs and opportunities. Sci. Total Environ. 601–602, 580–593. https://doi.org/10.1016/j.scitotenv.2017.05.212.

Liu, Y., Bralts, V.F., Engel, B.A., 2015. Evaluating the effectiveness of management practices on hydrology and water quality at watershed scale with a rainfall-runoff model. Sci. Total Environ. 511, 298–308. https://doi.org/ 10.1016/j.scitotenv.2014.12.077.

Liu, Y., Wang, Y., Sheng, H., Dong, F., Zou, R., Zhao, L., Guo, H., Zhu, X., He, B., 2014. Quantitative evaluation of lake eutrophication responses under alternative water diversion scenarios: A water quality modeling based statistical analysis approach. Sci. Total Environ. 468–469, 219–227. https://doi.org/10.1016/j.scitotenv.2013.08.054.

Loewen, M.R., Ackerman, J.D., Hamblin, P.F., 2007. Environmental implications of stratification and turbulent mixing in a shallow lake basin. Can. J. Fish. Aquat. Sci. 64, 43–57. https://doi.org/10.1139/f06-165.

Logan, T.J., Verhoff, F.H., DePinto, J.V., 1979. Biological Availability of Total Phosphorus. Lake Erie Wastewater Management Study. Technical Report Series. U.S. Army Corps of Engineers. Buffalo, N.Y.

Long, T., Wellen, C., Arhonditsis, G., Boyd, D., Mohamed, M., O'Connor, K., 2015. Estimation of tributary total phosphorus loads to Hamilton Harbour, Ontario, Canada, using a series of regression equations. J. Great Lakes Res. 41, 780–793. https://doi.org/10.1016/j.jglr.2015.04.001.

Long, T., Wellen, C., Arhonditsis, G., Boyd, D., 2014. Evaluation of stormwater and snowmelt inputs, land use and seasonality on nutrient dynamics in the watersheds of Hamilton Harbour, Ontario, Canada. J. Great Lakes Res. 40, 964–979. https://doi.org/10.1016/j.jglr.2014.09.017.

Lyons, J.E., Runge, M.C., Laskowski, H.P., Kendall, W.L., 2010. Monitoring in the context of structured decision-making and adaptive management. J. Wildl. Manag. 72, 1683–1692.

Maccoux, M.J., Dove, A., Backus, S.M., Dolan, D.M., 2016. Total and soluble reactive phosphorus loadings to Lake Erie: A detailed accounting by year, basin, country, and tributary. J. Great Lakes Res. 42, 1151–1165. https://doi.org/10.1016/j.jglr.2016.08.005.

Malkin, S.Y., Guildlford, S.J., Hecky, R.E., 2008. Modeling the growth response of Cladophora in a Laurentian Great Lake to the exotic invader Dreissena and to lake warming. Limnol. Oceanogr. 53(3), 1111–1124. https://doi.org/ 10.4319/lo.2008.53.3.1111.

Matisoff, G., Kaltenberg, E.M., Steely, R.L., Hummel, S.K., Seo, J., Gibbons, K.J., Bridgeman, T.B., Seo, Y., Behbahani, M., James, W.F., Johnson, L.T., Doan, P., Dittrich, M., Evans, M.A., Chaffin, J.D., 2016. Internal loading of phosphorus in western Lake Erie. J. Great Lakes Res. 42, 775–788. https://doi.org/10.1016/j.jglr.2016.04.004.

Matisoff, G., Neeson, T.M., 2005. Oxygen concentration and demand in Lake Erie sediments. J. Great Lakes Res. 31, 284–295. https://doi.org/10.1016/S0380-1330(05)70321-X.

May, L., Spears, B.M., 2012. Managing ecosystem services at Loch Leven, Scotland, UK: Actions, impacts and unintended consequences. Hydrobiologia. 681, 117–130. https://doi.org/10.1007/s10750-011-0931-x.

McDowell, R.W., Dils, R.M., Collins, A.L., Flahive, K., Sharpley, A.N., 2016. A review of the policies and implementation of practices to decrease water quality impairment by phosphorus in New Zealand, the UK, and the US. Nutr. Cycl. Agroecosyst. 104, 289–305.

McDonald, C.P., Urban, N.R., 2010. Using a model selection criterion to identify appropriate complexity in aquatic biogeochemical models. Ecol. Model. 221, 428–432. https://doi.org/10.1016/j.ecolmodel.2009.10.021.

McDonald, M.G., Harbaugh, A.W., 2003. The history of MODFLOW. Ground Water. 41, 280–283. https://doi.org/10.1111/j.1745-6584.2003.tb02591.x.

Meals, D.W., Dressing, S.A., Davenport, T.E. 2010. Lag time in water quality response to best management practices: a review. J. Environ. Qual. 39, 85–96. https://doi.org/10.2134/jeq2009.0108.

Mellina, E., Rasmussen, J.B., Mills, E.L., 1995. Impact of zebra mussel (Dreissena polymorpha) on phosphorus cycling and chlorophyll in lakes. Can. J. Fish. Aquat. Sci. 52, 2553–2573. https://doi.org/10.1139/f95-246.

Mellor, G.L., Yamada, T., 1982. Development of a turbulence closure model for geophysical fluid problems. Rev. Geophys. 20, 851–875. https://doi.org/10.1029/RG020i004p00851.

Michalak, A.M., Anderson, E.J., Beletsky, D., Boland, S., Bosch, N.S., Bridgeman, T.B., Chaffin, J.D., Cho, K., Confesor, R., Daloglu, I., DePinto, J. V., Evans, M.A., Fahnenstiel, G.L., He, L., Ho, J.C., Jenkins, L., Johengen, T.H., Kuo, K.C., LaPorte, E., Liu, X., McWilliams, M.R., Moore, M.R., Posselt, D.J., Richards, R.P., Scavia, D., Steiner, A.L., Verhamme, E., Wright, D.M., Zagorski, M.A., 2013. Record-setting algal bloom in Lake Erie caused by agricultural and meteorological trends consistent with expected future conditions. Proc. Natl. Acad. Sci. 110, 6448–6452. https://doi.org/10.1073/pnas.1216006110.

Millennium Ecosystem Assessment, 2005. Ecosystems and Human Well-being: Synthesis. Island Press, Washington, DC, pp.137.

Miller, J. E., 1984. Basic concepts of kinematic-wave models, U.S. Geol. Surv. Prof. Pap., 1302.

Millie, D.F., Fahnenstiel, G.L., Bressie, J.D., Pigg, R.J., Rediske, R.R., Klarer, D.M., Tester, P.A., and Litaker, R.W. 2009. Late-summer phytoplankton in western Lake Erie (Laurentian Great Lakes): bloom distributions, toxicity, and environmental influences. Aquat. Ecol. 43(4), 915–934. https://doi.org/10.1007/s10452-0099238-7.

Mitsch, W.J., Zhang, L., Stefanik, K.C., Nahlik, A.M., Anderson, C.J., Bernal, B., Hernandez, M., Song, K., 2012. Creating wetlands: primary succession, water quality changes, and self-design over 15 years. Bio. Sci. 3, 237–250.

Molot, L.A., Watson, S.B., Creed, I.F., Trick, C.G., McCabe, S.K., Verschoor, M.J., Sorichetti, R.J., Powe, C., Venkiteswaran, J.J., Schiff, S.L., 2014. A novel model for cyanobacterial bloom formation: the critical role of anoxia and ferrous iron. Freshw. Biol. 59, 1323–1340.

Mooij, W.M., Trolle, D., Jeppesen, E., Arhonditsis, G., Belolipetsky, P. V., Chitamwebwa, D.B.R., Degermendzhy, A.G., DeAngelis, D.L., De Senerpont Domis, L.N., Downing, A.S., Elliott, J.A., Fragoso, C.R., Gaedke, U., Genova, S.N., Gulati, R.D., Håkanson, L., Hamilton, D.P., Hipsey, M.R., 't Hoen, J., Hülsmann, S., Los, F.H., Makler-Pick, V., Petzoldt, T., Prokopkin, I.G., Rinke, K., Schep, S.A., Tominaga, K., van Dam, A.A., van Nes, E.H., Wells, S.A., Janse, J.H., 2010. Challenges and opportunities for integrating lake ecosystem modelling approaches. Aquat. Ecol. 44, 633–667. https://doi.org/10.1007/s10452-010-9339-3

Muenich, R.L., Kalcic, M., Scavia, D., 2016. Evaluating the impact of legacy P and agricultural conservation practices on nutrient loads from the Maumee River watershed. Environ. Sci. Technol. 50, 8146–8154. https://doi.org/10.1021/acs.est.6b01421.

Munawar, M., Munawar, I.F., Fitzpatrick, M., Niblock, H., Bower, K., Lorimer, J., 2008. An intensive assessment of planktonic communities in the Canadian waters of Lake Erie, 1998. In: Munawar, M., Heath, R. (Ed.), Checking the Pulse of Lake Erie. Michigan State University Press, East Lansing, Michigan, USA.

Murray, C., Sohngen, B., Pendleton, L., 2001. Valuing water quality advisories and beach amenities in the Great Lakes. Water Resour. Res. 37, 2583–2590. https://doi.org/10.1029/2001WR000409.

Murphy, D.V., Recous, S., Stockdale, E.A., Fillery, I.R.P., Jensen, L.S., Hatch, D.J., Goulding, K.W.T. 2003. Gross nitrogen fluxes in soil: Theory, measurement and application of 15N pool dilution techniques. In Sparks, D.L. (Ed.), Advances in Agronomy. Elsevier Science (USA), San Diego, California, USA.

Neitsch, S.L., Arnold,J.G., Kiniry,J.R., Williams,J.R., 2001. Soil and Water Assessment Tool Theoretical Documentation, Version 2000. Grassland, soil and water research service, Temple, TX.

Neuman, W.L., 2003. Social research methods: Qualitative and quantitative approaches. Fifth edition. Allyn and Bacon. Boston, Massachusetts, USA.

Neumann A., Dong F., Kim, D.-K, Shimoda, Y., Onandia, G., Javed, A., Arhonditsis, G.B., 2018. Development of an empirical watershed modelling framework to support adaptive management implementation. Ecol. Inform. In submission.

Obenour, D., Gronewold, A., Stow, C.A., Scavia, D., 2014. Using a Bayesian hierarchical model to improve Lake Erie cyanobacteria bloom forecasts. Water Resour. Res. 50, 7847–7860 https://doi.org/10.1002/2014WR015616.

Orr, P.T., Jones, G.J., 1998. Relationship between microcystin production and cell division rates in nitrogen-limited Microcystis aeruginosa cultures. Limnol. Oceanogr. 43, 1604–1614.

Osmond, D., Meals, D., Hoag, D., Arabi, M., Luloff, A., Jennings, G., McFarland, M., Spooner, J., Sharpley, A., Line, D., 2012. Improving conservation practices programming to protect water quality in agricultural watersheds: Lessons learned from the National Institute of Food and Agriculture–Conservation Effects Assessment Project. J. Soil. Water Conserv. 67, 122A–127A.

Oveisy, A., Rao, Y.R., Leon, L.F., Bocaniov, S.A., 2014. Three-dimensional winter modeling and the effects of ice cover on hydrodynamics, thermal structure and water quality in Lake Erie. J. Great Lakes Res. 40, 19–28. https://doi.org/ 10.1016/j.jglr.2014.09.008.

Oveisy, A., Boegman, L., Imberger, J., 2012. Three-dimensional simulation of lake and ice dynamics during winter. Limnol. Oceanogr. 57, 43–57. https://doi.org/10.4319/lo.2012.57.1.0043.

Ozersky, T., Malkin, S.Y., Barton, D.R., Hecky, R.E., 2009. Dreissenid phosphorus excretion can sustain C. glomerata growth along a portion of Lake Ontario shoreline. J Great Lakes Res. 35, 321–328.

Pahl-Wostl, C., 2009. A conceptual framework for analysing adaptive capacity and multi-level learning processes in resource governance regimes. Glob. Environ. Chang. 19, 354–365. https://doi.org/10.1016/j.gloenvcha. 2009.06.001.

Palm-Forster, L.H., Lupi, F., Chen, M., 2016. Valuing Lake Erie beaches using value and function transfers. Agric. Resour. Econ. Rev. 45, 270–292. https://doi.org/10.1017/age.2016.15.

Pappenberger, F., Beven, K.J., 2006. Ignorance is bliss: Or seven reasons not to use uncertainty analysis. Water Resour. Res. 42, W05302. https://doi.org/10.1029/2005WR004820.

Pascual, U., Muradian, R., Rodríguez, L.C., Duraiappah, A., 2010. Exploring the links between equity and efficiency in payments for environmental services: A conceptual approach. Ecol. Econ. 69, 1237–1244.

Paudel, R., Jawitz, J.W., 2012. Does increased model complexity improve description of phosphorus dynamics in a large treatment wetland? Ecol. Eng. 42, 283–294. https://doi.org/10.1016/j.ecoleng.2012.02.014.

Paytan, A., Roberts, K., Watson, S., Peek, S., Chuang, P.C., Defforey, D., Kendall, C., 2017. Internal loading of phosphate in Lake Erie central basin. Sci. Total Environ. 579, 1356–1365. https://doi.org/10.1016/j.scitotenv.2016.11.133.

Powers, S.M., Bruulsema, T.W., Burt, T.P., Chan, N.I., Elser, J.J., Haygarth, P.M., Howden, N.J.K., Jarvie, H.P., Lyu, Y., Peterson, H.M., Sharpley, A.N., Shen, J., Worrall, F., Zhang, F., 2016. Long-term accumulation and transport of anthropogenic phosphorus in three river basins. Nat. Geosci. 9, 353–356. https://doi.org/10.1038/ngeo2693.

Raftery, A.E., Li, N., Sevcikova, H., Gerland, P., Heilig, G.K., 2012. Bayesian probabilistic population projections for all countries. Proc. Natl. Acad. Sci. 50, 777–801. https://doi.org/10.1073/pnas.1211452109.

Ramin, M., Labencki, T., Boyd, D., Trolle, D., Arhonditsis, G.B., 2012a. A Bayesian synthesis of predictions from different models for setting water quality criteria. Ecol. Model. 242, 127–145. https://doi.org/10.1016/j.ecolmodel.2012.05.023.

Ramin, M., Perhar, G., Shimoda, Y., Arhonditsis, G.B., 2012b. Examination of the effects of nutrient regeneration mechanisms on plankton dynamics using aquatic biogeochemical modeling. Ecol. Model. 240, 139-155. http://dx.doi.org/10.1016/j.ecolmodel.2012.04.018

Ramin, M., Stremilov, S., Labencki, T., Gudimov, A., Boyd, D., Arhonditsis, G.B., 2011. Integration of numerical modeling and Bayesian analysis for setting water quality criteria in Hamilton Harbour, Ontario, Canada. Environ. Model. Softw. 147, 298–311. https://doi.org/10.1016/j.envsoft.2010.08.006.

Reavie, E.D., Cai, M., Twiss, M.R., Carrick, H.J., Davis, T.W., Johengen, T.H., Gossiaux, D., Smith, D.E., Palladino, D., Burtner, A., Sgro, G.V., 2016. Winter-spring diatom production in Lake Erie is an important driver of summer hypoxia. J. Great Lakes Res. 42, 608–618.

Refsgaard, J.C., Storm, B., 1995. MIKE SHE. In: Singh, V.P. (Ed.), Computer Models of Watershed Hydrology. Water Resources Pub., Highlands Ranch, Colorado, USA.

Richards, L.A., 1931. Capillary conduction of liquids through porous mediums. J. Appl. Phys. 1, 318–333. https://doi.org/10.1063/1.1745010.

Richards, P.R., Alameddine, I., David Allan, J., Baker, D.B., Bosch, N.S., Confesor, R., Depinto, J. V., Dolan, D.M., Reutter, J.M., Scavia, D., 2013. “Nutrient inputs to the Laurentian Great

Lakes by source and watershed estimated using SPARROW watershed models" by Dale M. Robertson and David A. Saad. J. Am. Water Resour. Assoc. 49, 715–724. https://doi.org/10.1111/jawr.12006.

Rinta-Kanto, J.M., Konopko, E.A., DeBruyn, J.M., Bourbonniere, R.A., Boyer, G.L., Wilhelm, S.W., 2009. Lake Erie Microcystis: Relationship between microcystin production, dynamics of genotypes and environmental parameters in a large lake. Harmful Algae. 8, 665–673. https://doi.org/10.1016/j.hal.2008.12.004.

Robertson, D.M., Saad, D.A., 2013. SPARROW models used to understand nutrient sources in the Mississippi/Atchafalaya River basin. J. Environ. Qual. 42, 1422–1440. https://doi.org/10.2134/jeq2013.02.0066.

Robertson, D.M., Saad, D.A., 2013. Reply to discussion - "Nutrient inputs to the Laurentian Great Lakes by source and watershed estimated using SPARROW watershed models." J. Am. Water Resour. Assoc. https://doi.org/10.1111/jawr.12060.

Rode, M., Wade, A.J., Cohen, M.J., Hensley, R.T., Bowes, M.J., Kirchner, J.W., Arhonditsis, G.B., Jordan, P., Kronvang, B., Halliday, S.J., Skeffington, R.A., Rozemeijer, J.C., Aubert, A.H., Rinke, K., Jomaa, S., 2016. Sensors in the stream: The high-frequency wave of the present. Environ. Sci. Technol. 50, 10297–10307. https://doi.org/10.1021/acs.est.6b02155.

Rossman, L.A., Huber, W., 2016. StormWater Management Model Reference Manual Volume III - Water Quality. US EPA Office of Research and Development, Washington, DC EPA/600/R-16/093. Available at: https://nepis.epa.gov/Exe/ZyPDF.cgi/P100P2NY.PDF?Dockey=P100P2NY.PDF.

Roulston, M.S., Smith, L.A., 2003. Combining dynamical and statistical ensembles. Tellus, Ser. A Dyn. Meteorol. Oceanogr. 55, 16–30. https://doi.org/10.1034/j.1600-0870.2003.201378.x.

Roy, E.D., Martin, J.F., Irwin, E.G., Conroy, J.D., Culver, D.A., 2011. Living within dynamic social-ecological freshwater systems: System parameters and the role of ecological engineering. Ecol. Eng. 37, 1661–1672. https://doi.org/10.1016/j.ecoleng.2011.06.044.

Roy, E.D., Martin, J.F., Irwin, E.G., Conroy, J.D., Culver, D.A., 2010. Transient social-ecological stability: The effects of invasive species and ecosystem restoration on nutrient management compromise in Lake Erie. Ecol. Soc. 15, 20.

Rucinski, D.K., DePinto, J. V., Beletsky, D., Scavia, D., 2016. Modeling hypoxia in the central basin of Lake Erie under potential phosphorus load reduction scenarios. J. Great Lakes Res. 42, 1206–1211. https://doi.org/10.1016/j.jglr.2016.07.001.

Rucinski, D.K., DePinto, J. V., Scavia, D., Beletsky, D., 2014. Modeling Lake Erie's hypoxia response to nutrient loads and physical variability. J. Great Lakes Res. 40, Supplement 3, 151–161. https://doi.org/10.1016/j.jglr.2014.02.003.

Rucinski, D.K., Beletsky, D., DePinto, J. V., Schwab, D.J., Scavia, D., 2010. A simple 1-dimensional, climate based dissolved oxygen model for the central basin of Lake Erie. J. Great Lakes Res. 36, 465–476. https://doi.org/10.1016/j.jglr.2010.06.002.

Scavia, D., Kalcic, M., Muenich, R.L., Read, J., Aloysius, N., Bertani, I., Boles, C., Confesor, R., DePinto, J., Gildow, M. and Martin, J., 2017. Multiple models guide strategies for agricultural nutrient reductions. Front. Ecol. Environ. 15(3), 126–132.

Scavia, D., DePinto, J.V., Bertani, I., 2016a. A multi-model approach to evaluating target phosphorus loads for Lake Erie. J. Great Lakes Res. 42(6), 1139–1150.

Scavia, D., DePinto, J.V., Auer, M., Bertani, I., Bocaniov, S., Chapra, S., Leon, L., McCrimmon, C., Obenour, D., Peterson, G., Rucinski, D., Schlea, D., Steger, C., Stumpf, R., Yerubandi, R., Zhang, H., 2016b. Great Lakes Water Quality Agreement Nutrient Annex Objectives and Targets Task Team Ensemble Multi-Modeling Report. Great Lakes National Program Office, USEPA, Chicago. Available at: https://www.epa.gov/sites/production/files/2016-09/documents/nutrientannex4multimodelingreportfinal31aug2016.pdf.

Scavia, D., Kalcic, M., Muenich, R.L., Aloysius, N., Arnold, J., Boles, C., Confessor, R., De Pinto, J., Gildow, M. and Martin, J., 2016c. Informing Lake Erie agriculture nutrient management via scenario evaluation. University of Michigan: Ann Arbor, MI, USA.

Scavia, D., David Allan, J., Arend, K.K., Bartell, S., Beletsky, D., Bosch, N.S., Brandt, S.B., Briland, R.D., Daloğlu, I., DePinto, J. V., Dolan, D.M., Evans, M.A., Farmer, T.M., Goto, D., Han, H., Höök, T.O., Knight, R., Ludsin, S.A., Mason, D., Michalak, A.M., Peter Richards, R., Roberts, J.J., Rucinski, D.K., Rutherford, E., Schwab, D.J., Sesterhenn, T.M., Zhang, H., Zhou, Y., 2014. Assessing and addressing the re-eutrophication of Lake Erie: Central basin hypoxia. J. Great Lakes Res. 40, 226–246. https://doi.org/10.1016/j.jglr.2014.02.004

Schloesser, D.W., Stickel, R.G., Bridgeman, T.B., 2005. Potential oxygen demand of sediments from Lake Erie. J. Great Lakes Res. 31, 272–283. https://doi.org/10.1016/S0380-1330(05)70320-8.

Schwab, D.J., Beletsky, D., DePinto, J., and Dolan, D.M. 2009. A hydrodynamic approach to modeling phosphorus distribution in Lake Erie. J. Great Lakes Res., 35(1), 50–60. https://doi.org/10.1016/J.Jglr.2008.09.003.

Schwarz, G.E., Hoos, A.B., Alexander, R.B., Smith, R.A., 2006. The SPARROW Surface Water-Quality Model: Theory, Applications and User Documentation. U.S. Geol. Surv. Tech. Methods.

Sharpley, A.N., Kleinman, P.J., Flaten, D.N. and Buda, A.R., 2011. Critical source area management of agricultural phosphorus: experiences, challenges and opportunities. Water Sci. Technol. 64(4), 945–952.

Sharpley, A.N., Daniel, T., Gibson, G., Bundy, L., Cabrera, M., Sims, T., Stevens, R., Lemunyon, J., Kleinman, P., Parr, R., 2006. Best Management Practices To Minimize Agricultural Phosphorus Impacts on Water Quality. U.S. Department of Agriculture, Agricultural Research Service. ARS-163 (50 pp).

Sharpley, A.N. and Williams, J.R., 1990. EPIC Erosion/Productivity Impact Calculator: 1. Model Documentation. USA Department of Agriculture Technical Bulletin No. 1768, USA Government Printing Office, Washington DC.

Shimoda, Y., Onandia, G., Brady, M., Xu, Z., Arhonditsis, G.B., 2018. Dissecting the ecological foundation of a multi-model ensemble: What have we learned for the mechanisms that drive eutrophication in Lake Erie? Ecol. Inform. In submission.

Shimoda, Y., Azim, M.E., Perhar, G., Ramin, M., Kenney, M.A., Sadraddini, S., Gudimov, A. and Arhonditsis, G.B., 2011. Our current understanding of lake ecosystem response to climate change: What have we really learned from the north temperate deep lakes? J. Great Lakes Res. 37(1), 173–193.

Shimoda, Y., Arhonditsis, G.B., 2015. Integrating hierarchical Bayes with phosphorus loading modelling. Ecol. Inform. 29, 77–91. https://doi.org/10.1016/j.ecoinf.2015.07.005.

Shimoda, Y., Arhonditsis, G.B., 2016. Phytoplankton functional type modelling: Running before we can walk? A critical evaluation of the current state of knowledge. Ecol. Model. 320, 29–43. https://doi.org/10.1016/j.ecolmodel.2015.08.029.

Shimoda, Y., Watson, S.B., Palmer, M.E., Koops, M.A., Mugalingam, S., Morley, A., Arhonditsis, G.B., 2016a. Delineation of the role of nutrient variability and dreissenids (Mollusca, Bivalvia) on phytoplankton dynamics in the Bay of Quinte, Ontario, Canada. Harmful Algae. 55, 121–136. https://doi.org/10.1016/j.hal.2016.02.005.

Shimoda, Y., Rao, Y.R., Watson, S. and Arhonditsis, G.B., 2016b. Optimizing the complexity of phytoplankton functional group modeling: An allometric approach. Ecol. Inf. 31, 1–17.

Shuchman, R., Korosov, A., Hatt, C., Pozdnyakov, D., Means, J., Meadows, G., 2006. Verification and application of a bio-optical algorithm for Lake Michigan using SeaWiFS: A 7-year inter-annual analysis. J. Great Lakes Res. 32, 258–279. https://doi.org/10.3394/0380-1330(2006)32[258:VAAOAB]2.0.CO;2.

Smith, D.A., Matisoff, G., 2008. Sediment oxygen demand in the central basin of Lake Erie. J. Great Lakes Res. 34, 731–744.

Smith, D.R., Wilson, R.S., King, K.W., Zwonitzer, M., McGrath, J.M., Harmel, R.D., Haney, R.L. and Johnson, L.T., 2018. Lake Erie, phosphorus, and microcystin: Is it really the farmer's fault? J. Soil Water Conserv. 73(1), 48–57.

Sohngen, B., King, K.W., Howard, G., Newton, J., Forster, D.L., 2015. Nutrient prices and concentrations in Midwestern agricultural watersheds. Ecol. Econ. 112, 141–149.

Spillman, C.M., Hamilton, D.P., Hipsey, M.R., Imberger, J., 2008. A spatially resolved model of seasonal variations in phytoplankton and clam (Tapes philippinarum) biomass in Barbamarco Lagoon, Italy. Estuar. Coast. Shelf Sci. 79, 187–203. https://doi.org/10.1016/j.ecss.2008.03.020.

Steffen, M.M., Belisle, B.S., Watson, S.B., Boyer, G.L. and Wilhelm, S.W., 2014. Status, causes and controls of cyanobacterial blooms in Lake Erie. J. Great Lakes Res. 40(2), 215–225.

Steinman, A.D., Cardinale, B.J., Munns, W.R., Ogdahl, M.E., Allan, J.D., Angadi, T., Bartlett, S., Brauman, K., Byappanahalli, M., Doss, M., Dupont, D., Johns, A., Kashian, D., Lupi, F., McIntyre, P., Miller, T., Moore, M., Muenich, R.L., Poudel, R., Price, J., Provencher, B., Rea, A., Read, J., Renzetti, S., Sohngen, B., Washburn, E., 2017. Ecosystem services in the Great Lakes. J. Great Lakes Res. 43, 161–168. https://doi.org/10.1016/j.jglr.2017.02.004.

Stow, C.A., Cha, Y., Johnson, L.T., Confesor, R., Richards, R.P., 2015. Long-term and seasonal trend decomposition of Maumee River nutrient inputs to western Lake Erie. Environ. Sci. Technol. 49, 3392–3400. https://doi.org/10.1021/es5062648.

Stumpf, R.P., Wynne, T.T., Baker, D.B. and Fahnenstiel, G.L., 2012. Interannual variability of cyanobacterial blooms in Lake Erie. PloS one, 7(8), p.e42444.

Stumpf, R.P., Johnson, L.T., Wynne, T.T., Baker, D.B., 2016. Forecasting annual cyanobacterial bloom biomass to inform management decisions in Lake Erie. J. Great Lakes Res. 42 (6), 1174–1183.

Tayfur, G., Kavvas, M.L., Govindaraju, R.S., Storm, D.E., 1993. Applicability of St. Venant equations for two-dimensional overland flows over rough infiltrating surfaces. J. Hydraul. Eng-ASCE. 119, 51–63.

Teubner, K., Tolotti, M., Greisberger, S., Morscheid, H., Dokulil, M.T. and Morscheid, H., 2003. Steady state phytoplankton in a deep pre-alpine lake: species and pigments of epilimnetic versus metalimnetic assemblages. In Naselli-Flores, L., Padisák, J., Bach, M.F. (Ed.), Phytoplankton and Equilibrium Concept: The Ecology of Steady-State Assemblages (pp. 49-64). Springer, Dordrecht.

Tomlinson, L., Auer, M., Bootsma, H., 2010. The Great Lakes Cladophora model: development and application to Lake Michigan. J. Great Lakes Res. 36, 287–297.

Tuholske, J., and Kilbert, K., 2015. Moving Forward: Legal solutions to Lake Erie's harmful algal blooms. Lucas County. Lucas County Government.

Twiss, M.R., McKay, R.M.L., Bourbonniere, R.A., Bullerjahn, G.S., Carrick, H.J., Smith, R.E.H., Winter, J.G., D'souza, N.A., Furey, P.C., Lashaway, A.R., Saxton, M.A., Wilhelm, S.W., 2012. Diatoms abound in ice-covered Lake Erie: an investigation of offshore winter limnology in Lake Erie over the period 2007 to 2010. J. Great Lakes Res. 38, 18–30.

Valipour, R., Leon, L., Depew, D., Dove, A., Rao, Y., 2016. High-resolution modeling for development of nearshore ecosystem objectives in eastern Lake Erie. J. Great Lakes Res. 42 (6), 1241–1251.

Vanderploeg, H.A., Johengen, T.H., Liebig, J.R., 2009. Feedback between zebra mussel selective feeding and algal composition affects mussel condition: Did the regime changer pay a price for its success? Freshw. Biol. 54, 47–63. https://doi.org/10.1111/j.1365-2427.2008.02091.x.

Vanderploeg, H.A., Liebig, J.R., Carmichael, W.W., Agy, M.A., Johengen, T.H., Fahnenstiel, G.L., Nalepa, T.F., 2001. Zebra mussel (Dreissena polymorpha) selective filtration promoted toxic Microcystis blooms in Saginaw Bay (Lake Huron) and Lake Erie. Can. J. Fish. Aquat. Sci. 58, 1208–1221. https://doi.org/10.1139/cjfas-58-6-1208.

Vanni M.J., Flecker, A.S., Hood, J.M., Headworth, J.L., 2002. Stoichiometry of nutrient cycling by vertebrates in a tropical stream: linking species identity and ecosystem processes. Ecol. Lett. 5, 285–93.

Verhamme, E.M., Redder, T.M., Schlea, D.A., Grush, J., Bratton, J.F., DePinto, J. V., 2016. Development of the Western Lake Erie Ecosystem Model (WLEEM): Application to connect phosphorus loads to cyanobacteria biomass. J. Great Lakes Res. 42, 1193–1205. https://doi.org/10.1016/j.jglr.2016.09.006.

Walters, C.J., Holling, C.S., 1990. Large-scale management experiments and learning by doing. Ecology. 71, 2060–2068. https://doi.org/10.2307/1938620.

Walters, C.J., 1986. Adaptive management of renewable resources. Macmillan Publishers Ltd. New York, New York, USA.

Watson, S.B., Miller, C., Arhonditsis, G., Boyer, G.L., Carmichael, W., Charlton, M., Confessor, R., Depew, D.C., Höök, T.O., Ludsin, S., Matisoff, G., McElmurry, S.P., Murray, M.W., Richards, P., Rao, Y.R., Steffen, M., Wilhelm, S. Harmful algal blooms and the re-eutrophication of Lake Erie: past, present and future management scenarios. Harmful Algae 56, 44–66.

Wellen, C., Kamran-Disfani, A.R., Arhonditsis, G.B., 2015. Evaluation of the current state of distributed watershed nutrient water quality modeling. Environ. Sci. Technol. 49, 3278–3290. https://doi.org/10.1021/es5049557.

Wellen, C., Arhonditsis, G.B., Long, T., Boyd, D., 2014. Accommodating environmental thresholds and extreme events in hydrological models: A Bayesian approach. J. Great Lakes Res. 40, Supplement 3, 102–116. https://doi.org/10.1016/j.jglr.2014.04.002.

Wellen, C., Arhonditsis, G.B., Long, T., Boyd, D., 2014. Quantifying the uncertainty of nonpoint source attribution in distributed water quality models: A Bayesian assessment of SWAT's sediment export predictions. J. Hydrol. 519, 3353–3368. https://doi.org/10.1016/j.jhydrol.2014.10.007.

Wellen, C., Arhonditsis, G.B., Labencki, T., Boyd, D., 2012. A Bayesian methodological framework for accommodating interannual variability of nutrient loading with the SPARROW model. Water Resour. Res. 48, W10505. https://doi.org/10.1029/2012WR011821.

Wilks, D.S., 2002. Realizations of daily weather in forecast seasonal climate. J. Hydrometeorol. 3(2), 195–207.

Williams, B.K., Brown, E.D., 2016. Technical challenges in the application of adaptive management. Biol. Conserv. 195, 255–263. https://doi.org/10.1016/j.biocon.2016.01.012.

Williams, B.K. and Johnson, F.A., 2015. Value of information and natural resources decision-making. Wildl. Soc. Bull. 39(3), 488–496.

Williams, B.K. and Brown, E.D., 2014. Adaptive management: from more talk to real action.J. Environ. Manage.53(2), 465–479.

Williams, B.K., Eaton, M.J. and Breininger, D.R., 2011. Adaptive resource management and the value of information. Ecol. Model. 222(18), 3429–3436.

Williams, B.K., Szaro, R.C. and Shapiro, C.D., 2009. Adaptive management: the US Department of the Interior technical guide. US Department of the Interior.

Wilson, K.A., Howell, E.T., Jackson, D.A., 2006. Replacement of zebra mussels by quagga mussels in the Canadian nearshore of Lake Ontario: the importance of substrate, round goby abundance, and upwelling frequency. J. Great Lakes Res. 32, 11–28.

Wolf, D., Georgic, W., Klaiber, H.A., 2017. Reeling in the damages: Harmful algal blooms' impact on Lake Erie's recreational fishing industry. J. Environ. Manage. 199, 148–157. https://doi.org/10.1016/j.jenvman.2017.05.031.

Wynne, T.T., Stumpf, R.P., Tomlinson, M.C., Dyble, J., 2010. Characterizing a cyanobacterial bloom in western Lake Erie using satellite imagery and meteorological data. Limnol. Oceanogr. 55, 2025–2036. https://doi.org/10.4319/lo.2010.55.5.2025.

Wynne, T.T., Stumpf, R.P., Tomlinson, M.C., Warner, R.A., Tester, P.A., Dyble, J., Fahnenstiel, G.L., 2008. Relating spectral shape to cyanobacterial blooms in the Laurentian Great Lakes. Int. J. Remote Sens. 29, 3665–3672. https://doi.org/10.1080/01431160802007640.

Xue, Y., David, M.B., Gentry, L.E. and Kovacic, D.A., 1998. Kinetics and modeling of dissolved phosphorus export from a tile-drained agricultural watershed. J. Environ. Qual. 27(4), 917–922.

Yaksich, S.M., Melfi, D.A., Baker, D.B., Kramer, J.W., 1985. Lake Erie nutrient loads, 1970–1980. J. Great Lakes Res. 11, 117–131.

Yokota, F., Thompson, K.M., 2004. Value of information analysis in environmental health risk management decisions: past, present, and future. Risk Anal. 24, 635–650. https://doi.org/10.1111/j.0272-4332.2004.00464.x.

Zehe, E., Sivapalan, M., 2009. Threshold behaviour in hydrological systems as (human) geo-ecosystems: Manifestations, controls, implications. Hydrol. Earth Syst. Sci. 13, 1273–1297. https://doi.org/10.5194/hess-13-1273-2009.

Zehe, E., Maurer, T., Ihringer, J., Plate, E., 2001. Modeling water flow and mass transport in a loess catchment. Phys. Chem. Earth, Part B Hydrol. Ocean. Atmos. 26, 487–507. https://doi.org/10.1016/S1464-1909(01)00041-7.

Zhang, H., Culver, D.a., Boegman, L., 2008. A two-dimensional ecological model of Lake Erie: application to estimate dreissenid impacts on large lake plankton populations. Ecol. Model. 214, 219–241. https://doi.org/10.1016/j.ecolmodel.2008.02.005.

Zhang, H., Culver, D., Boegman, L., 2016. Spatial distributions of external and internal phosphorus loads and their impacts on Lake Erie's phytoplankton. J. Great Lakes Res. 42 (6), 1212–1227.

Zhao, Y., Jones, M.L., Shuter, B.J. and Roseman, E.F., 2009. A biophysical model of Lake Erie walleye (Sander vitreus) explains interannual variations in recruitment. Can. J. Fish. Aquat. Sci. 66(1), 114–125.

Zhou, Y., Obenour, D.R., Scavia, D., Johengen, T.H., Michalak, A.M., 2013. Spatial and temporal trends in Lake Erie hypoxia, 1987–2007. Environ. Sci. Technol. 47, 899–905.

Zolfaghari, K., Duguay, C., 2016. Estimation of Water Quality Parameters in Lake Erie from MERIS Using Linear Mixed Effect Models. Remote Sens. 8(6), 473. https://doi.org/10.3390/rs8060473.

8 Electronic Supplement

Additional electronic tables and graphs have been submitted as separate documents. Brief descriptions of contents are listed below.

List of Figures

effectiveness for *SRP* and *TP* for filter strips and conservation tillage (Gitau et al., 2005), where negative values indicate that the *BMP* acts as a *P* source; and *(d)* and *(e)* illustrate the probability distributions of *BMP* effectiveness on *SRP* and *TP* reduction for reduced tillage and wetland restoration, respectively (Igras, 2016).

List of Tables

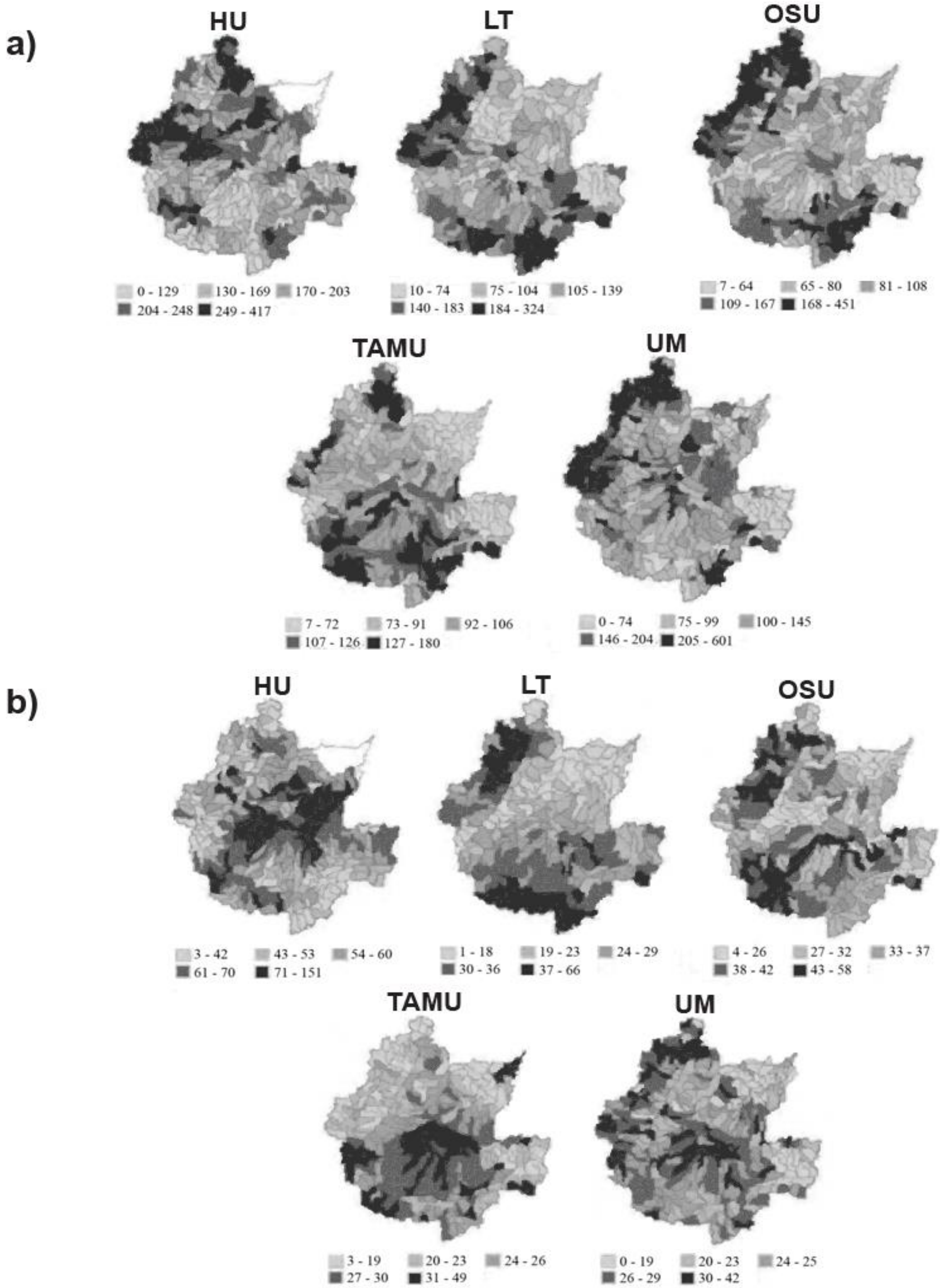

a)
HU
0 - 129
130 - 169
170 - 203
204 - 248
249 - 417
LT
10 - 74
75 - 104
105 - 139
140 - 183
184 - 324
OSU
7 - 64
65 - 80
81 - 108
109 - 167
168 - 451
TAMU
7 - 72
73 - 91
92 - 106
107 - 126
127 - 180
UM
0 - 74
75 - 99
100 - 145
146 - 204
205 - 601
b)
HU
3 - 42
43 - 53
54 - 60
61 - 70
71 - 151
LT
1 - 18
19 - 23
24 - 29
30 - 36
37 - 66
OSU
4 - 26
27 - 32
33 - 37
38 - 42
43 - 58
TAMU
3 - 19
20 - 23
24 - 26
27 - 30
31 - 49
UM
0 - 19
20 - 23
24 - 25
26 - 29
30 - 42

Figure S1

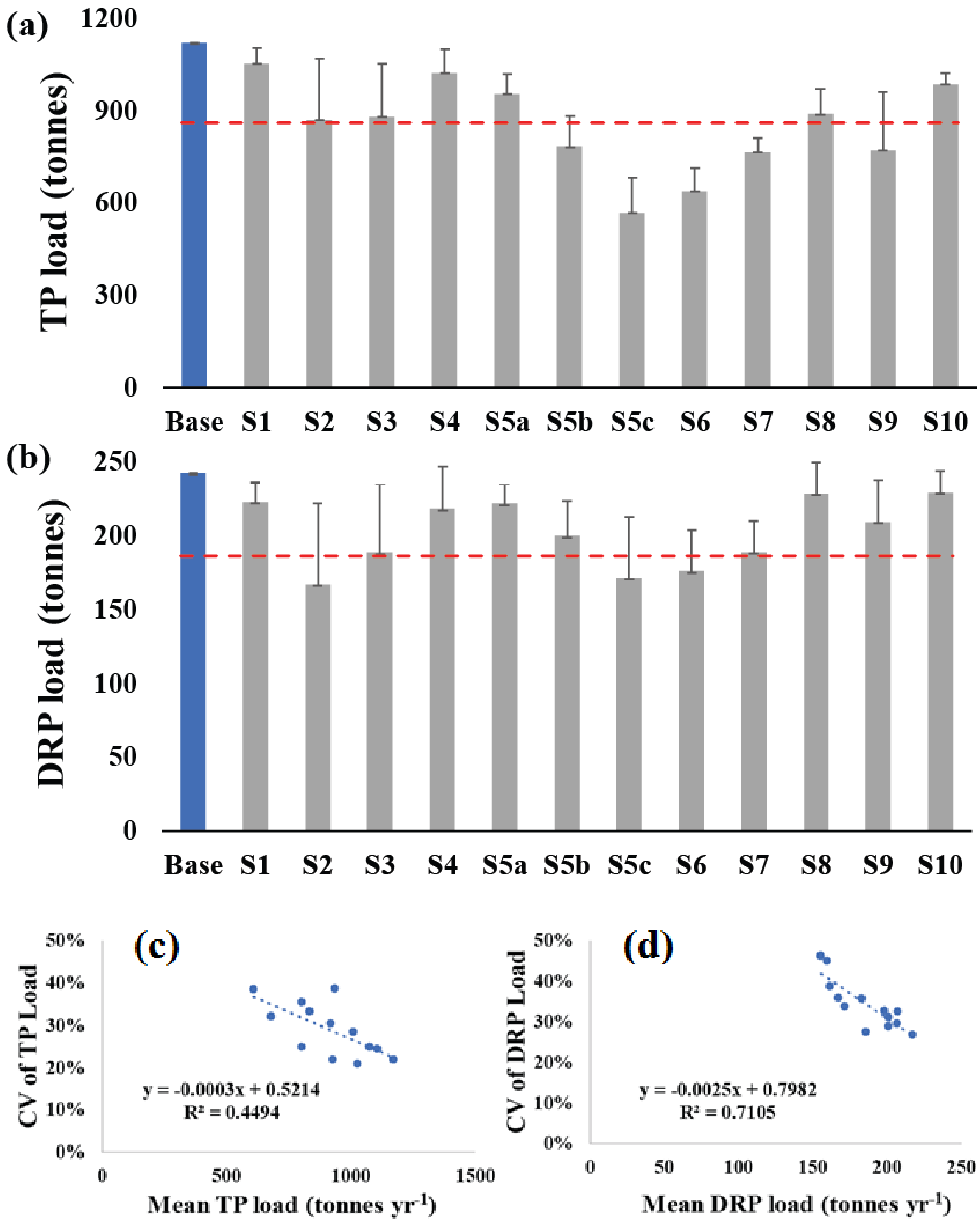


Figure S2

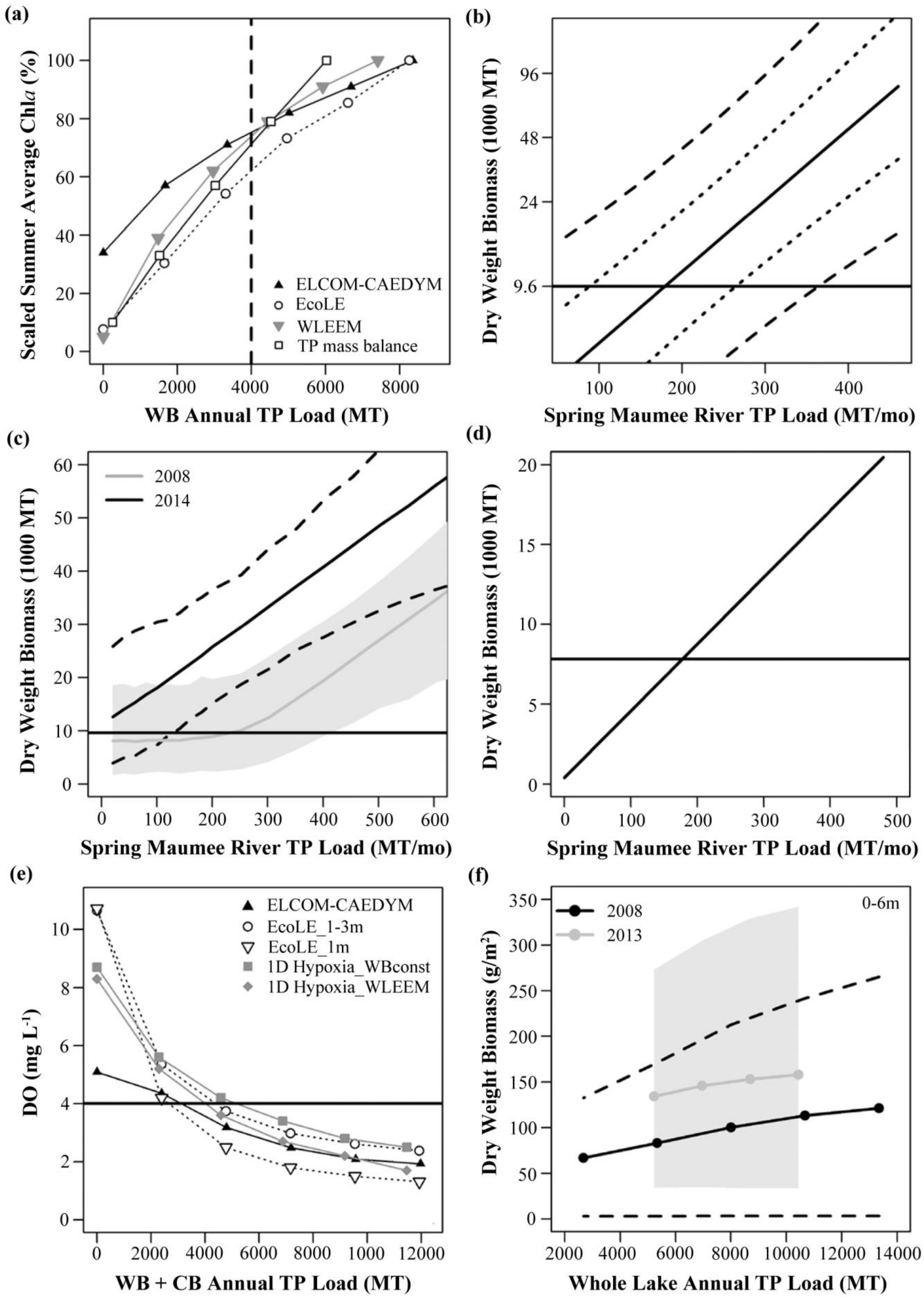

(a)
Scaled Summer Average Chla (%)
WB Annual TP Load (MT)
ELCOM-CAEDYM
EcoLE
WLEEM
TP mass balance
(b)
Dry Weight Biomass (1000 MT)
Spring Maumee River TP Load (MT/mo)
(c)
2008
2014
Dry Weight Biomass (1000 MT)
Spring Maumee River TP Load (MT/mo)
(d)
Dry Weight Biomass (1000 MT)
Spring Maumee River TP Load (MT/mo)
(e)
ELCOM-CAEDYM
EcoLE_1-3m
EcoLE_1m
1D Hypoxia_WBconst
1D Hypoxia_WLEEM
DO (mg L-1)
WB + CB Annual TP Load (MT)
(f)
2008
2013
0-6m
Dry Weight Biomass (g/m2)
Whole Lake Annual TP Load (MT)

Figure S3

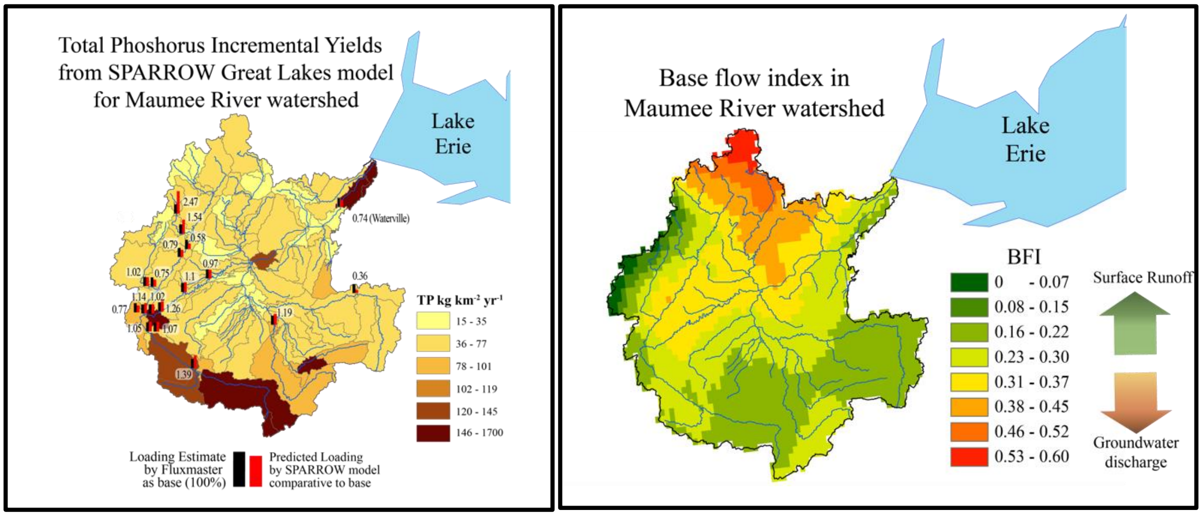


Figure S4

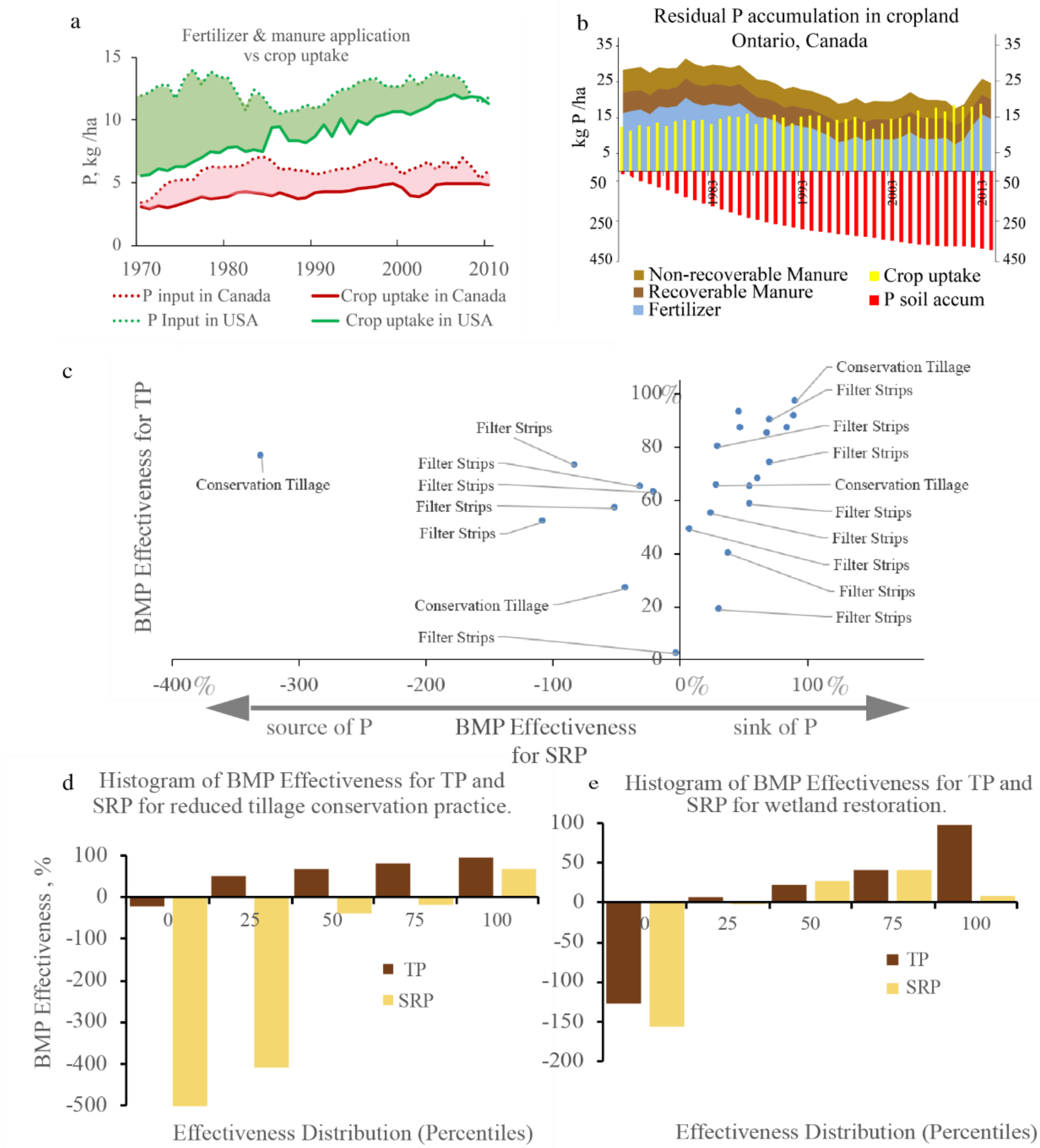

a
Fertilizer & manure application
vs crop uptake
P, kg /ha
1970
1980
1990
2000
2010
P input in Canada
Crop uptake in Canada
P Input in USA
Crop uptake in USA
b
Residual P accumulation in cropland
Ontario, Canada
kg P /ha
Non-recoverable Manure
Recoverable Manure
Fertilizer
Crop uptake
P soil accum
c
BMP Effectiveness for TP
Conservation Tillage
Filter Strips
source of P
BMP Effectiveness
for SRP
sink of P
d
Histogram of BMP Effectiveness for TP and
SRP for reduced tillage conservation practice.
BMP Effectiveness , %
TP
SRP
Effectiveness Distribution (Percentiles)
e
Histogram of BMP Effectiveness for TP and
SRP for wetland restoration.
TP
SRP
Effectiveness Distribution (Percentiles)

Figure S5

Table S1: Description of the different model scenarios, categorized based on the policy questions they were intended to address (modified from Scavia et al., 2016).

Number of Scenarios	Name	Description	Policy question
S1	**Nutrient management** at 25% random adoption	25% of row crop acreage was randomly chosen and the following practices were applied: a 50% reduction in P fertilizer application, fall timing of P applications, and subsurface placement of P into the soil.	What level of nutrient management will be sufficient to reach phosphorus targets?
S2	**Nutrient management** at 100% adoption	The following practices were applied to 100% of row crop fields throughout the watershed: a 50% reduction in P fertilizer application, fall timing of P applications, and subsurface placement of P into the soil.	Can nutrient management alone achieve targets?
S3	C**ommonly recommended practices** at 100% random adoption	Four practices were applied to a separate 25% of the crop fields: a 50% reduction in P fertilizer application, subsurface application of P fertilizers, continuous no tillage, and medium-quality buffer strips.	What extent of adoption of commonly recommended practices will be needed to achieve the targets?
S4	**Continuous no-tillage and subsurface placement of P fertilizer** at 50% random adoption	50% of row crop acres were randomly chosen to apply a combination of continuous no-tillage and subsurface application of P fertilizers.	Is no-tillage effective provided P is applied below the soil surfaces?
S5	**Cropland conversion to grassland** at 10% (5a), 25% (5b), and 50% (5c) targeted adoption	In this set of three scenarios, 10%, 25%, and 50% of the row croplands with the lowest crop yields and greatest TP losses were converted to switchgrass and managed for wildlife habitat with limited harvesting for forage and no P fertilization.	How much row cropland would be needed to be converted to grassland to achieve the targets?
S6	**Series of practices** at 50% targeted adoption	A series of practices were targeted to the 50% of row cropland with the highest TP loss in the watershed. Practices: subsurface application of P fertilizers, cereal rye cover crop in the winters without wheat, and application of medium-quality buffer strips.	What extent of targeted infield and edge-of-field practices reach the targets?

Table S1 (Continued): Description of the different model scenarios, categorized based on the policy questions they were intended to address.

Number of Scenarios	Name	Description	Policy question
S7	**Series of practices** at 50% random adoption	A series of practices were applied at random to 50% of row cropland. Practices: subsurface application of P fertilizers, cereal rye cover crop in the winters without wheat, and application of medium-quality buffer strips.	What if in-field and edge-of-field practices were applied at random?
S8	**Diversified rotation** at 50% random adoption	An alternative corn-soybean-wheat rotation with a cereal rye cover crops all winters without wheat was applied over a randomly chosen 50% of row cropland.	What is the impact of returning to winter wheat and winter cover crops?
S9	**Wetlands and buffer strips** at 25% targeted adoption	Wetlands were targeted to the 25% of sub-watersheds with the greatest P loading and assumed to intercept half of overland and tile flow, and medium-quality buffer strips were targeted to the 25% of row crop acreage responsible for the greatest TP loss.	How much P reduction can be achieved through structural practices?
S10	**In-field practices** at 25% random adoption	A group of four in-field practices was applied to a random 25% of row cropland. Practices included a 50% reduction in P fertilizer application, fall timing of P applications, subsurface placement of P fertilizers, and a cereal rye cover crop.	What can be achieved at 25% application of in-field practices?

Table S2: Knowledge gaps and sources of uncertainty to guide monitoring and improve modelling in Lake Erie.

Knowledge gaps	
Phytoplankton	• Winter productivity under ice needs to be further examined to establish the causal linkages with spring phytoplankton dynamics and summer hypoxic conditions.
Cyanobacteria /HABs	• Increased *SRP* loading since mid-1990s appears to correlate with more frequent and severe *HAB*s. Accurate fractionation of exogenous phosphorus loads (i.e., bioavailability) is needed to confirm the relationship with algal blooms. • Regular monitoring of cyanotoxin data (Microcystin-*LR*) in the nearshore zone. • Quantitative characterization of the selective rejection of cyanobacteria by dreissenids
Hypoxia	• It is critical to understand the intensive microbiological, geochemical, and physical processes occurring within the top few centimeters of the sediment and determine the fraction of organic matter and nutrients released into the overlying water. Field, experimental (e.g., porewater analysis, phosphorus fractionation, organic matter profiles), and modelling (e.g., primary and secondary redox reactions, mineral precipitation dissolution reactions, acid dissociation reactions, and P binding form reactions) work should be designed to shed light on the mechanisms of organic matter mobilization in the sediments and to identify process controls under a variety of conditions.
Dreissenids	• Empirical information on selective rejection of cyanobacteria by dreissenids and revisit the mathematical representation of dreissenid feeding accordingly. • Empirical data on the spatial distribution of zebra and quagga mussels is needed, since the two species differ significantly with regard to filtration and excretion rates. • Empirical data on the year-to-year variability of dreissenid population density and size distribution is critical to realistically predict their impact on water quality. • The role of dreissenids on *N* mineralization and the modification of the *N:P* ratios remains understudied
Cladophora	• Local modelling efforts in Lake Erie will greatly benefit from a high-resolution monitoring of the nearshore zone to establish the causal linkages between the abiotic conditions (e.g., phosphate, light, temperature) in the surrounding environment and the internal *P* content and sloughing rates in *Cladophora* mats.

Sediment-related processes	• Empirical information is needed to distinguish the relative importance of oxic and anoxic release of P from the sediments in the western basin of Lake Erie in order to validate model predictions. • Empirical information is needed to quantify the contribution of particulate matter resuspension to the *P* concentrations in the water column (all models except *WLEEM* do not consider this process separately from *P* chemical release).
Refinement of monitoring framework	
	• Monitoring programs should be extended until the end of September to obtain a better characterization of hypoxia, given that the maximum hypoxic area in the nearshore zone has been projected to occur in September (Bocaniov and Scavia, 2016). • Monitoring programs should target nearshore areas, which occupy a significant portion of the area of the central basin and represent an important habitat for many aquatic species as well as a source of drinking water (Bocaniov and Scavia, 2016). • Whole-year phytoplankton and zooplankton sampling to quantitatively characterize the seasonal succession patterns, as well as the likelihood of top-down control.

Table S3: Research on ecosystem service valuation concentrated on water quality improvements, erosion risk reduction, recreation, and recreational fishing in Lake Erie.

Method	***Data requirements***	***Advantages***	***Notes / Limitations***	***Examples in Lake Erie***
Hedonic valuation (revealed preference): assumes that the value individuals place on a commodity is based on the attributes it possesses.	Detailed information on: • Property market prices • Individual property characteristics • Distance from environmental attributes	• Based on observed behaviour in property values	• Property market should be near-equilibrium with an appropriate size and coverage • Assumes that buyers have complete knowledge about environmental attributes • Difficult to isolate the effects of the environmental attribute on property value	Dorfman et al. (1996); Kriesel et al. (1993)
Travel-cost valuation (revealed preference): rationalizes that the value of a site is reflected in the willingness-to-pay the associated travel cost.	Detailed information on: • Individual travel costs • On-site expenses • License fees (if applicable) • Capital expenditure on recreational equipment (if applicable) • Value of time spent travelling • Socioeconomic characteristics of users	• Based on observed behaviour	• Travel distances are ideally relatively short, and sample has a variety of distances, costs and socioeconomic characteristics • Can be confounded if the visit is not intended for the specific ecosystems service (i.e., trip made for multiple destinations) • Assumes that users have complete knowledge about the ecosystem service being used • Cannot value ecosystems that are not visible or well understood (e.g., nutrient cycling, erosion control)	Chen (2013); Hushak et al. (1988); Kelch et al. (2006); Murray et al. (2001); Sohngen et al. (1999); Sohngen et al. (2015)
Contingent valuation (stated preference): uses a context of a hypothetical market in which individuals report about their willingness-to-pay for ecosystem services (for which markets do not exist) through questionnaires and/or interviews.	Detailed information on: • Ecosystem characteristics/function • Socioeconomic characteristics of respondents	• Can be used to value almost all environmental attributes (i.e., ecosystem services without markets or parallel markets)	• Tend to be cost and time intensive to implement • Biases are common in survey responses due to hypothetical nature of the market • Low income constrains willingness-to-pay for ecosystem services	Brox et al. (1996); Kreutzwiser (1981)
Benefit value transfer (benefits transfer): estimates values by transferring values from a	• Primary valuation results for a site similar to target site	• More cost and time effective than primary valuation studies • Recommended to be used to scope if a more in-depth primary valuation study is required	• Estimates from primary studies can be outdated • Results tend to be less accurate than primary studies since estimates are unlikely to be perfectly transferable	Hayder (2014); Palm-Forster et al.

primary valuation study conducted for a similar site.				(2016); Wolf et al. (2017)
Benefit function transfer (benefits transfer): adjusts for differences in the characteristics of the population/site between the primary study and target sites. The result is more relevant to the targeted site.	• Primary valuation results for a site similar to target site • Detailed site and population characteristics for both the primary study and target sites	• More cost and time effective than primary valuation studies (although requires more resources than benefit value transfers) • Recommended to be used to scope if a more in-depth primary valuation study is required	• Estimates from primary studies can be outdated • Results tend to be less accurate than primary studies since estimates are unlikely to be perfectly transferable (although result is more robust than benefit value transfer)	Palm-Forster et al. (2016)

References

Bocaniov, S.A., Scavia, D., 2016. Temporal and spatial dynamics of large lake hypoxia: Integrating statistical and three-dimensional dynamic models to enhance lake management criteria. Water Resour. Res. 52, 4247–4263. https://doi.org/10.1002/2015WR018170

Bouwman, L., Goldewijk, K.K., Van Der Hoek, K.W., Beusen, A.H.W., Van Vuuren, D.P., Willems, J., Rufino, M.C., Stehfest, E., 2013. Exploring global changes in nitrogen and phosphorus cycles in agriculture induced by livestock production over the 1900-2050 period. Proc. Natl. Acad. Sci. 110, 20882–20887. https://doi.org/10.1073/pnas.1012878108

Brox, J.A., Kumar, R.C., Stollery, K.R., 1996. Willingness To Pay for Water Quality and Supply Enhancements in the Grand River Watershed. Can. Water Resour. J. 21, 275–288. https://doi.org/10.4296/cwrj2103275

Chen, M., 2013. Valuation of Public Great Lakes Beaches in Michigan. Michigan State University.

Dorfman, J.H., Keeler, A.G., Kriesel, W., 1996. Valuing Risk-Reducing Interventions with Hedonic Models: The Case of Erosion Protection. J. Agric. Resour. Econ. 21, 109–119.

Gitau, M.W., Gburek, W.J., Jarrett, A.R., 2005. A tool for estimating best management practice effectiveness for phosphorus pollution control. J. Soil Water Conserv. 60, 1–10.

Hayder, S., 2014. Socio-Economic Impact of the Presence of Asian Carp in the Great Lakes Basin. Winnipeg, Manitoba.

Hushak, L.J., Winslow, J.M., Dutta, N., 1988. Economic value of Great Lakes sportfishing: The case of private-boat fishing in Ohio’s Lake Erie. Trans. Am. Fish. Soc. 117, 363–373. https://doi.org/10.1577/1548-8659(1988)117<0363:EVOGLS>2.3.CO;2

Igras, J.D.M., 2016. Uncertainty Analysis of the Performance of a System of Best Management Practices for Achieving Phosphorus Load Reduction to Surface Waters. The University of Western Ontario.

International Plant Nutrition Institute, Canada Balance 2013, Crop Nutrient Balance, 2014, http://phosphorus.ipni.net/beagle/PPP-3033&f=Canada%20Balance%202013.xlsx

Kelch, D., Lichtkoppler, F., Sohngen, B., Daigneault, A., 2006. The Value of Steelhead (Onchorhynchus mykiss) Angling in Lake Erie Tributaries. J. Great Lakes Res. 32, 424–433. https://doi.org/10.3394/0380-1330(2006)32

Kreutzwiser, R., 1981. The Economic Significance of the Long Point Marsh, Lake Erie, as a Recreational Resource. J. Great Lakes Res. 7, 105–110. https://doi.org/10.1016/S0380-1330(81)72034-3

Kriesel, W., Randall, A., Lichtkoppler, F., 1993. Estimating the Benefits of Shore Erosion Protection in Ohio's Lake Erie Housing Market. Water Resour. Res. 29, 795–801.

Murray, C., Sohngen, B., Pendleton, L., 2001. Valuing water quality advisories and beach amenities in the Great Lakes. Water Resour. Res. 37, 2583–2590. https://doi.org/10.1029/2001WR000409

Palm-Forster, L.H., Lupi, F., Chen, M., 2016. Valuing Lake Erie beaches using value and function transfers. Agric. Resour. Econ. Rev. 45, 270–292. https://doi.org/10.1017/age.2016.15

Scavia, D., Kalcic, M., Muenich, R.L., Aloysius, N., Arnold, J., Boles, C., Confesor, R., DePinto, J., Gildow, M., Martin, J., Read, J., Redder, T., Robertson, D.M., Sowa, S.P., Wang, Y.-C., White, M., Yen, H., 2016a. Informing Lake Erie agriculture nutrient management via scenario evaluation. Ann Arbor, Michigan.

Scavia, D., DePinto, J.V., Bertani, I., 2016b. A multi-model approach to evaluating target phosphorus loads for Lake Erie. J. Great Lakes Res. 42(6), 1139–1150.

Sohngen, B., Lichtkoppler, F., Bielen, M., 1999. The Value of Day Trips to Lake Erie Beaches. Columbus, Ohio. https://doi.org/Technical Bulletin Series, Publication OSHU-TB-039

Sohngen, B., Zhang, W., Bruskotter, J., Sheldon, B., 2015. Results from a 2014 survey of Lake Erie anglers: Final report submitted to the Lake Erie Protection Fund. Columbus, Ohio. https://doi.org/10.13140/RG.2.1.3875.2084

Wolf, D., Georgic, W., Klaiber, H.A., 2017. Reeling in the damages: Harmful algal blooms' impact on Lake Erie's recreational fishing industry. J. Environ. Manage. 199, 148–157. https://doi.org/10.1016/j.jenvman.2017.05.031