Multiple models guide strategies for agricultural nutrient reductions

Donald Scavia[1]*, Margaret Kalcic[1,2], Rebecca Logsdon Muenich[1], Jennifer Read[1], Noel Aloysius[2], Isabella Bertani[1], Chelsie Boles[3], Remegio Confesor[4], Joseph DePinto[3], Marie Gildow[2], Jay Martin[2,8], Todd Redder[3], Dale Robertson[5], Scott Sowa[6], Yu-Chen Wang[1], and Haw Yen[7]

In response to degraded water quality, federal policy makers in the US and Canada called for a 40% reduction in phosphorus (P) loads to Lake Erie, and state and provincial policy makers in the Great Lakes region set a load-reduction target for the year 2025. Here, we configured five separate SWAT (US Department of Agriculture's Soil and Water Assessment Tool) models to assess load reduction strategies for the agriculturally dominated Maumee River watershed, the largest P source contributing to toxic algal blooms in Lake Erie. Although several potential pathways may achieve the target loads, our results show that any successful pathway will require large-scale implementation of multiple practices. For example, one successful pathway involved targeting 50% of row cropland that has the highest P loss in the watershed with a combination of three practices: subsurface application of P fertilizers, planting cereal rye as a winter cover crop, and installing buffer strips. Achieving these levels of implementation will require local, state/provincial, and federal agencies to collaborate with the private sector to set shared implementation goals and to demand innovation and honest assessments of water quality-related programs, policies, and partnerships.

Front Ecol Environ 2017; 15(3): 126–132, doi:10.1002/fee.1472

Many coastal marine and freshwater ecosystems across the US are increasingly exhibiting symptoms of eutrophication, most often caused by excess inputs of nitrogen (N) and phosphorus (P). Primary among these symptoms are increases in the extent and duration of harmful algal blooms (HABs; Paerl and Paul 2012) and depleted levels of dissolved oxygen (hypoxia; Diaz and Rosenberg 2008), and the resulting impacts range from contaminated drinking water to fish kills and loss of critical fish habitat. The primary, and in most cases the only, effective strategy for mitigating these effects is to reduce N and P inputs.

Although impacts such as these were once substantially reduced in the Laurentian Great Lakes, they have resurfaced, particularly in Lake Erie (Scavia *et al*. 2014a). For instance, under the 1978 binational Great Lakes Water Quality Agreement (GLWQA), reductions in point sources of P resulted in a 50% reduction in total phosphorus (TP) loading, with associated improvements in water quality and fisheries (Charlton *et al*. 1993; Ludsin *et al*. 2001). However, with changes in the ecology, climate, and the now dominant nonpoint P sources, Lake Erie HABs and hypoxia have increased markedly since the mid-1990s (Bridgeman *et al*. 2013; Scavia *et al*. 2014a). The hypoxic area is now often greater than 4000 km^2, with a record of 8800 km^2 set in 2012 (Zhou *et al*. 2015). Toxic *Microcystis* blooms set records in 2011 (Michalak *et al*. 2013) and again in 2015. Despite evidence of a potential role of N in late summer (Chaffin *et al*. 2013), development of the blooms has been strongly connected to P loads (Obenour *et al*. 2014; Scavia *et al*. 2014b).

In response to these changes, the US and Canada agreed to revise Lake Erie's loading targets (GLWQA 2012). To guide the new targets, a multi-model effort including both mechanistic and statistical models was used to generate load-response curves (Scavia *et al*. 2014b, 2016a). Based largely on information from this multi-model effort, new target loads were proposed (GLWQA 2015), ultimately established by the US and Canada, and supported by the region's governors and premiers, as well as by the International Joint Commission (a bi-national organization, established under the Boundary Waters Treaty of 1909, that is dedicated to water-resource conflict resolution). The new targets call for reducing annual and spring (March–July) P loads to Lake Erie by 40% from their 2008 levels (GLWQA 2016). The spring target loads for the Maumee River are 860 metric tons (MT) of TP and 186 MT of dissolved reactive phosphorus (DRP). Although the federal governments did not set a date by which the targets should be met, Michigan, Ohio, and the Canadian province of Ontario set one for 2025.

The task ahead is to implement programs to achieve that reduction, primarily from the now dominant and

---

[1] *Water Center, Graham Sustainability Institute, University of Michigan, Ann Arbor, MI* *(scavia@umich.edu); [2] *Department of Food, Agricultural, and Biological Engineering, The Ohio State University, Columbus, OH;* [3] *LimnoTech, Ann Arbor, MI;* [4] *National Center for Water Quality Research, Heidelberg University, Tiffin, OH;* [5] *Water Science Center, US Geological Survey (USGS), Middleton, WI;* [6] *The Nature Conservancy, Lansing, MI;* [7] *Blackland Research & Extension Center, Texas A&M University, Temple, TX;* [8] *Ohio Sea Grant, Columbus, OH*

harder to treat nonpoint sources – mainly from agriculture (IJC 2014). Here, we describe a new multi-model effort designed to inform the binational process for controlling these P sources. We focus on Lake Erie's western basin (WB) (Figure 1) because it is the site of the most extensive toxic cyanobacteria blooms and a prime source of nutrients driving central basin hypoxia (Bridgeman *et al.* 2013; Scavia *et al.* 2014b; Maccoux *et al.* 2016). The WB loads come overwhelmingly from the Maumee and Detroit rivers (Figure 1) at approximately equal loads. The Maumee River has a relatively low flow and high P concentrations, whereas the Detroit River has very high flow and relatively low P concentrations which are well below thresholds for producing cyanobacteria blooms. We therefore focus on the Maumee River where the vast majority of P is delivered from agricultural sources (Han *et al.* 2012; Scavia *et al.* 2016a). In addition, the 40 years of daily load estimates obtained from a gage station near the mouth of the Maumee River at Waterville, Ohio (Baker *et al.* 2014) provide an important check on the models and make the Maumee an excellent example for other agriculturally dominated watersheds.

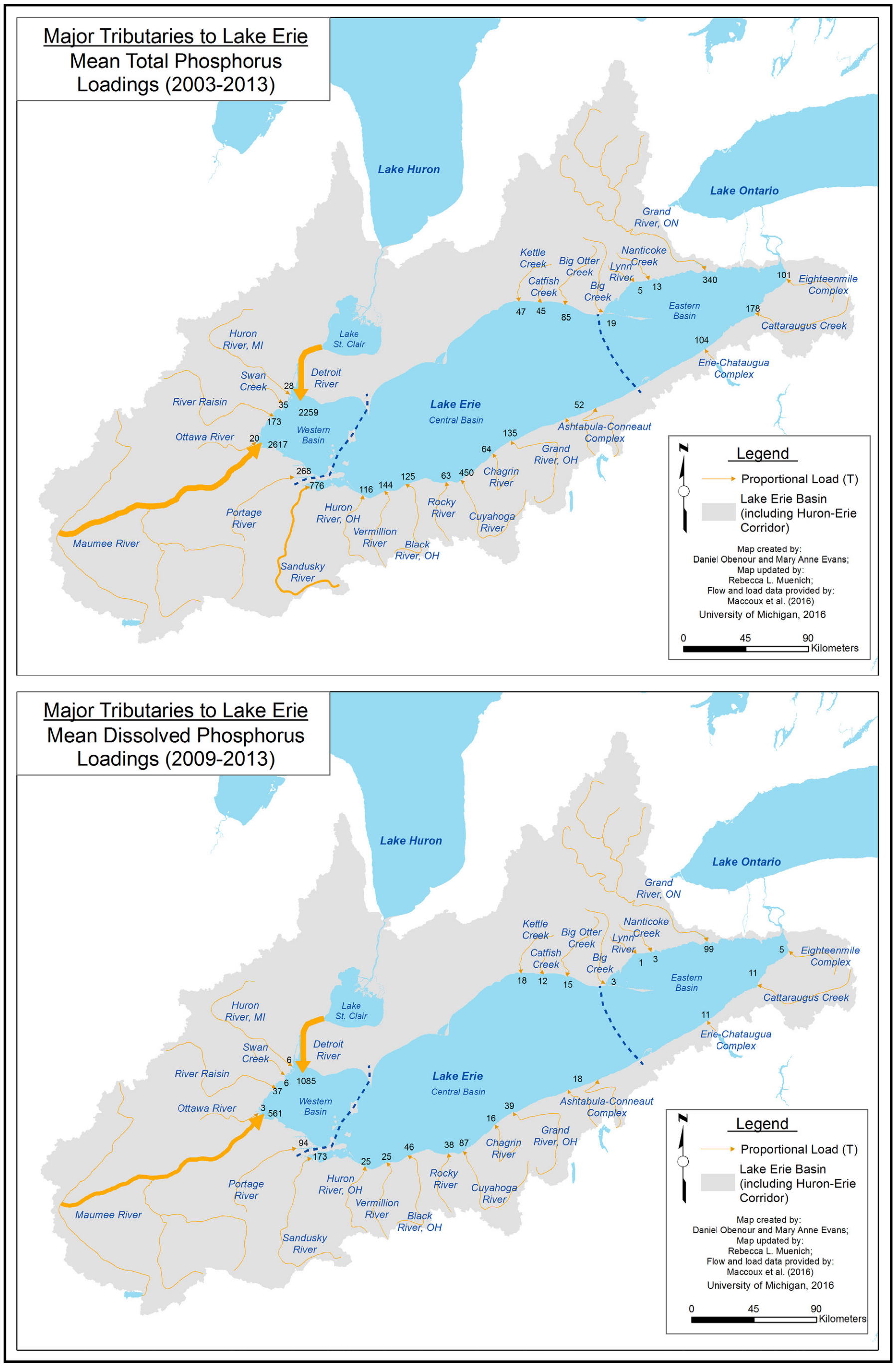

Figure 1. *Relative average annual phosphorus loads – total phosphorus from 2003 to 2013 (top panel) and dissolved phosphorus from 2009 to 2013 (bottom panel) – for the major tributaries of Lake Erie. Loads are proportional to the drawn river-arrow widths. Redrawn from Maccoux* et al. *(2016).*

Methods

A multi-model approach

The use of multiple models provides benefits that a single model cannot, including viewing problems from different conceptual and operational perspectives, using common datasets in different ways, providing multiple lines of evidence, and reducing decision risk based on a diversity of perspectives. Multi-model efforts for lakes and estuaries include those used in the 1970s (Bierman 1980) and more recently to establish target loads for the Great Lakes (Scavia *et al.* 2016a) and for managing nutrient loads to the Chesapeake Bay (Weller *et al.* 2013), the Gulf of Mexico (Scavia *et al.* 2004), and the Neuse River Estuary (Stow *et al.* 2003). Although ensemble modeling has been applied to evaluate and compare hydrological predictions (Breuer *et al.* 2009; Seiller *et al.* 2012; Velazquez *et al.* 2013), few studies have used it to assess watershed water quality (Boomer *et al.* 2013) and none have applied this approach to evaluate policy-relevant land management scenarios.

For this analysis, we assembled five models that rely on different implementations of the US Department of Agriculture's (USDA's) Soil and Water Assessment Tool

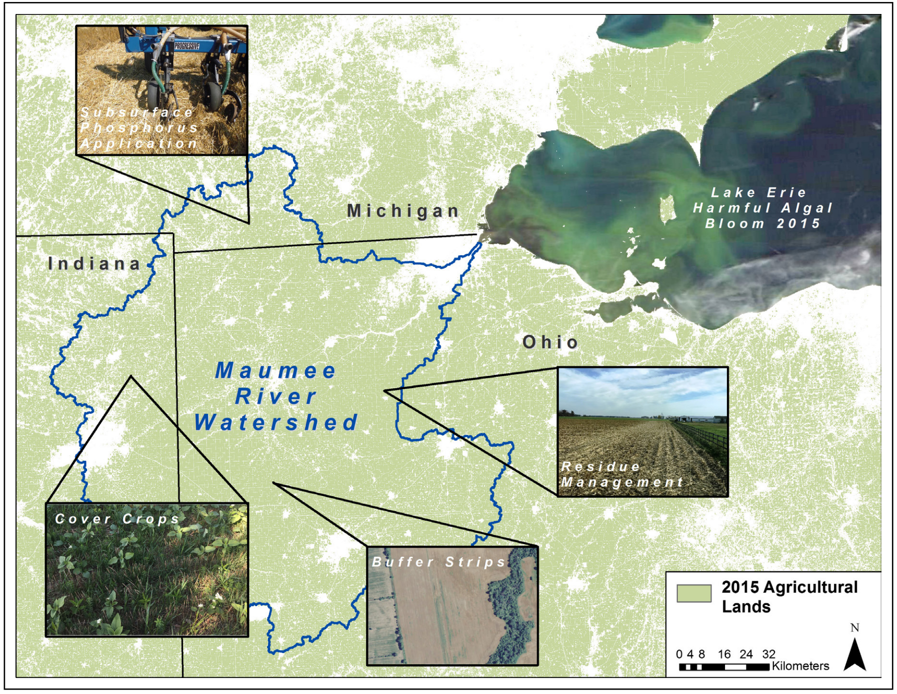

Figure 2. *Extent of agricultural lands in the WB of Lake Erie watersheds. Data sources: US land cover data from the 2015 USDA Cropland Data Layer; Canadian data from the 2015 AAFC Annual Crop Inventory; imagery of harmful algal bloom is a MODIS true-color image from 2 Aug 2015 retrieved online from NASA Worldview, courtesy of the NASA EOSDIS Land Processes Distributed Active Archive Center (LP DAAC), USGS/Earth Resources Observation and Science Center, Sioux Falls, SD. Credits (clockwise from top left): J Martin, Y-C Wang, USDA Farm Service Agency National Aerial Imagery Program 2014, and R Muenich.*

(SWAT; Arnold *et al.* 1998). Even though the models use the same base SWAT framework, they are actually quite different because of the independent critical decisions made about spatial resolution, data sources, subroutines, land management operations, and model parameterization and calibration approaches (see WebTables 1–3). While there may be a temptation to select one model based on "superior performance", there are many methods for evaluating performance and it is not easy to decide which model has the best fit for a complex watershed. For example, goodness-of-fit can be evaluated through various graphical and statistical methods for streamflow and loads, but it can also be assessed through measures of performance related to ensuring field-level nutrient export, soil nutrient content, and crop yields are within observed ranges (Yen *et al.* 2014). We chose to use multiple models because their accuracy in representing the baseline condition is not uniquely quantifiable and each model represents a reasonable representation of the real world.

Each of the five models was built and calibrated prior to this study. Although calibrated to various time periods between 1990 and 2012, the models were all validated for this effort to a common period (2005–2014) to serve as a baseline for comparisons. For validation, all models used the same precipitation, temperature, and point-source discharges. Model performance was evaluated by: (1) standard statistical tests (R^2, Nash-Sutcliffe efficiency [NSE]), percent bias (PBIAS), root mean squared error-observations standard deviation ratio (RSR) for monthly streamflow and P loads; (2) comparisons between time series and boxplots of modeled and measured streamflow and P loads; and (3) internal checks against estimates of percent of flow through subsurface drains and average crop yields. For more details on the differences among the model characteristics and their validation tests, see WebTables 1–4, WebFigure 1, and Scavia *et al.* (2016b).

Following validation, each model was used to analyze scenarios (Figure 2; Table 1) that were developed in consultation with representatives from environmental and agricultural communities, and guided by analyses of single-practice scenarios (Scavia *et al.* 2016b). These scenarios ranged from modest implementation of combinations of commonly applied practices (Scenarios 3–11) to extreme levels of implementation with those less commonly applied, such as converting row crops to switchgrass (Scenarios 2a–c). In addition to these agricultural conservation scenarios, for comparison we ran hypothetical extreme Scenario 1, which eliminates all point sources. Providing a range of modest to extreme, and even unrealistic, scenarios helps generate the information needed to inform decisions about the required levels of implementation. To compare with the targets, we multiplied the percent differences from each model's baseline condition by the average observed load at the Waterville gage station from 2005–2014. For ensemble model scenarios, we calculated a weighted average (with associated 95% confidence intervals) across models' predictions. Weights (WebTable 5) were developed with a Bayesian model averaging framework by estimating each posterior probability of being correct given the observed data, thereby reflecting the model's predictive performance over the validation dataset (Raftery *et al.* 2005; Duan *et al.* 2007).

Results and discussion

Although some of the validation statistics among the models (WebTable 4) were better than others for certain variables (eg TP versus DRP versus flow), all models were judged to be suitable for inclusion within

Table 1. Brief descriptions of the scenarios and results (more details are provided in WebTable 6)

No.	*Scenario description*
1	All point source discharges removed
2a-c	10, 25, 50% of lowest yielding row croplands with greatest P losses converted to switchgrass for wildlife habitat
3	50% less P fertilizer, fall application, subsurface placement, and cover crops applied together on random 25% of row cropland
4	50% less P fertilizer, fall application, subsurface placement applied together on random 25% of row cropland
5	Same as #4, but applied to 100% of row cropland
6	50% less P fertilizer, subsurface application, no-till, medium-quality buffers each applied to separate 25% of row cropland
7	No-till and subsurface P application applied together on a randomly selected 50% of row cropland
8	Subsurface P application, cereal rye cover crop in the winters without wheat, and medium-quality buffer strips on 50% of row cropland with the highest P loss
9	Same as #8 but applied to a random 50% of row cropland
10	A corn–soybean–wheat rotation with a cereal rye cover crop all winters without wheat applied randomly on 50% of row cropland
11	Wetlands targeted to 25% of sub-watersheds with greatest P loss, and buffer strips targeted to the 25% of row cropland responsible for the greatest P loss

Note: P: phosphorus.

the ensemble based on common criteria (Engel *et al.* 2007; Moriasi *et al.* 2007). There is variability among models, but the weighted average of all five models compares well with observations (WebFigure 1), especially for TP. The models tend toward slight overprediction of DRP at low load levels.

Given the difference in model assumptions, applications, and development, there was reassuring consistency among the models in estimating responses to the different scenarios (Figure 3; WebTable 6; for individual model results, see Scavia *et al.* [2016b]). All scenarios resulted in lowering TP and DRP loads, with load reductions increasing with greater scale of implementation and targeting areas of high P loss. Extreme Scenario 1, which eliminates all point-source discharges, reduced the March–July TP and DRP loads by only 5% and 10%, respectively, illustrating the importance of nonpoint sources. The land conversion scenarios (2a–c) are rather extreme and are unlikely to be implemented. They were included to illustrate that 25–50% of land would have to be removed from production to achieve the target loads if no additional nutrient management and in-field or edge-of-field practices were employed. For all other scenarios, the impact on total crop production was minor (WebTable 7).

The most promising scenarios included widespread use of nutrient management practices (especially subsurface application of P fertilizers, as also seen in single-practice scenarios; Scavia *et al.* 2016b) and installation of buffer strips. Because the US Farm Bill limits data access, we were not able to identify the extent or location of many existing practices for the models. Practices that might already be in place but were not captured in the baseline models include buffer strips, cover crops aside from winter wheat, and wetlands. For these practices, the best interpretation of our results is that they identify the need for *additional implementation*. For instance, to achieve a result similar to Scenario 9, an *additional* 50% of cereal rye and buffer strips are required. Other existing practices – such as fall timing of P applications, subsurface placement of P, continuous no-tillage, winter wheat grown in rotation, and reduced fertilizer application rates – are included to some extent in the baseline models. The best interpretation of our results for these practices, as well as land conversation to switchgrass, is that they identify the required *total level of implementation*. NRCS (2016) estimated that 99% of cropland in the WB watershed has at least one conservation practice in place, but it is clear from these results that more widespread adoption of additional, more effective practices is required to meet the targeted reductions, a conclusion also drawn in the USDA's Conservation Effects Assessment Project for the same region (chapter 5 in NRCS 2016).

Our results suggest there are several pathways to achieve the new target loads for Lake Erie. However, all the successful pathways require broad implementation of both common and less common practices. For example, three scenarios that appear to reach the TP goal (Figure 3a) tested targeted (Scenario 8) or random (Scenario 9) treatment of 50% of croplands in combination with nutrient management, cover crops, and buffer strips, or a combination of wetlands and buffer strips on 25% of cropland or sub-basins, respectively (Scenario 11). A comparison of Scenarios 8 and 9 highlights the importance of carrying out mitigating practices in areas where they are needed most, ie where P loads are most critical. Identifying these specific locations was beyond the scope of this work, but can be accomplished in consultation with conservationists

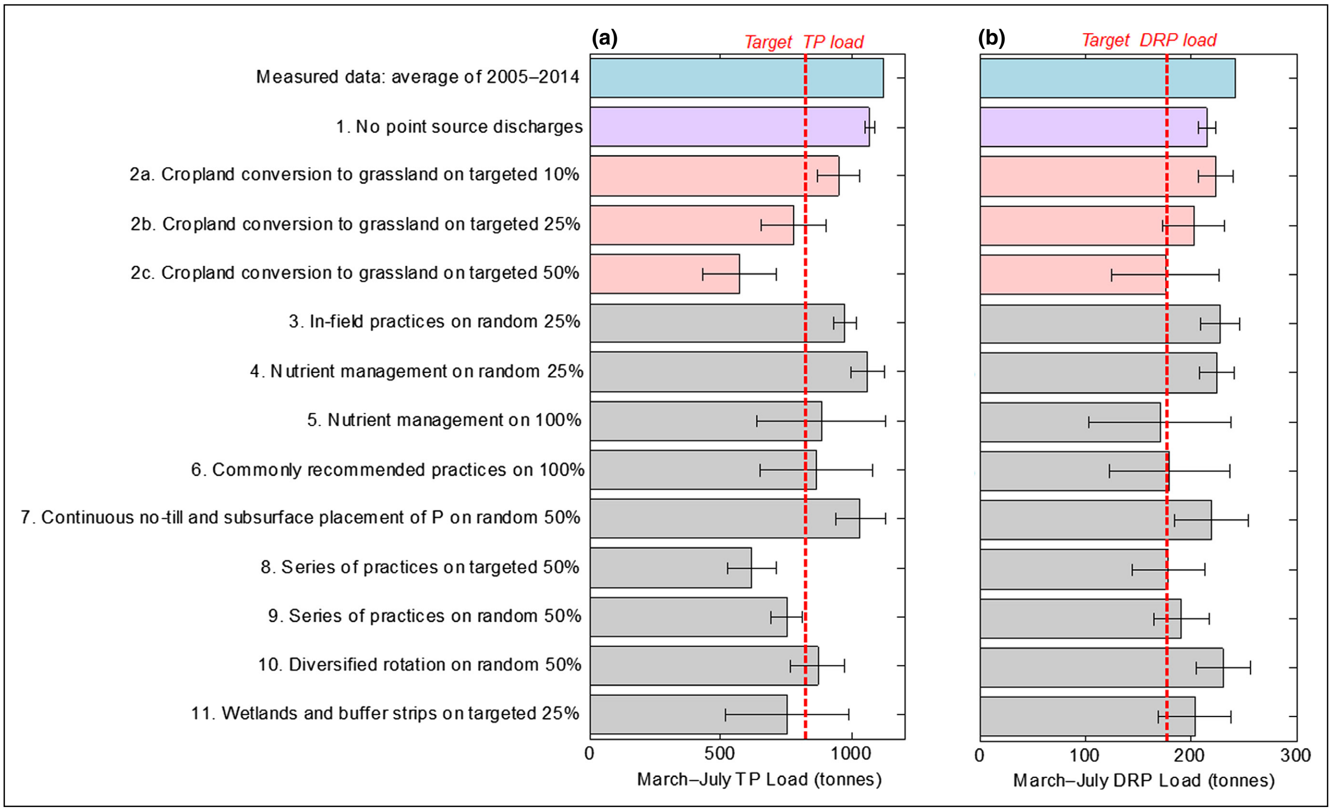

Figure 3. *Weighted average and 95% confidence intervals of the five SWAT models' March–July TP (a) and DRP (b) loads during the 2005–2014 modeling time period. The average observed March–July loads (area-weighted to Waterville, Ohio gage station) from 2005 to 2014 are represented in the top bars and the GLWQA target loads are depicted by the dashed red lines. Scenario 1 is the result of removing all point-source discharges; Scenarios 2a–c show a dose response as to how much land would need to be converted to grassland in order to meet the targets without going beyond current agricultural conservation measures; Scenarios 3–11 demonstrate the effect of implementing more agricultural conservation. DRP = dissolved reactive phosphorus; GLWQA = Great Lakes Water Quality Agreement; P = phosphorus; SWAT = Soil and Water Assessment Tool; TP = total phosphorus.*

and producers who have intimate knowledge of farm landscapes. Two scenarios – 5 (reduced P rates, fall P application, and subsurface placement of P on 100% of fields) and 2c (targeted conversion of 50% of cropland to grassland) – achieved the DRP target loads (Figure 3b). Scenario 8 (targeted series of practices, mentioned above) was very close to meeting the DRP target. Scenario 5 highlights the importance of the proper rate and correct placement of P applications promoted by the Natural Resources Conservation Service (NRCS) 590 standard and 4R nutrient management practices (Bruulsema *et al.* 2009). A 4R certification program in the WB, launched in 2014, certified nutrient management plans on 26% of the cropland in the basin in 2 years (Vollmer-Sanders *et al.* 2016). Scenario 5 also produced TP reductions near the 40% goal. These results illustrate that, while substantial rates and coverage of implementation will be required, there are several pathways for achieving the new loading targets for both TP and DRP.

Not all potential practices or combinations of practices were simulated in this work; however, reaching the new target loads is clearly a daunting task and will require extensive changes in management and much greater investment of resources to achieve the required levels of implementation, particularly for the less commonly applied practices. These findings are consistent with those of other recent studies that assessed management scenarios needed to achieve water quality and biological goals for streams in the Saginaw Bay, MI watershed (Sowa *et al.* 2016) and in the WB (chapter 5 in NRCS 2016; Kalcic *et al.* 2016; Keitzer *et al.* 2016; Muenich *et al.* 2016). Results across those studies also show that funding within the conservation provisions of the current US Farm Bill alone is insufficient to address these problems. Additional and targeted funding for the most effective conservation practices is needed, but other approaches may also be required, such as implementing conservation-compliance policies that target the most effective practices for a given field, funding only the most effective practices, and developing market-shaping policies related to modifying human dietary choices and altering energy production to reduce the demand for crops responsible for high P loss.

Key agencies at the local, state/provincial, and federal levels must join with the private sector and use the information from these studies to help set and achieve water-quality goals as well as to assess existing and potential conservation-oriented practices, programs, policies, and partnerships. Fortunately, innovative efforts like water funds, pay-for-performance, and public–private partnerships are being implemented within the WB and other parts of the Great Lakes region and hold promise (Fales *et al.* 2016). NRCS's current 3-year US$41-million investment to target, expand, and accelerate conservation practices in the WB is an important step in the right direction. The challenge is how to integrate and scale up these new approaches so they treat the number of acres needed to realize measureable improvements in water quality. However, at this time, it is not clear whether current programs have sufficient funding or policies in place that enable targeting of the best practices to the right places at the right scales to support the level of implementation needed to achieve the 40% load reduction targets.

Lake Erie is just one of many watersheds faced with the difficult task of finding sustainable solutions to reducing nonpoint source impacts on socially valued ecosystem services like freshwater provision, recreation, and fishing. Many other watersheds clearly demonstrate the difficulty in addressing nonpoint source pollution and meeting water quality goals. For instance, the goal of reducing the Gulf of Mexico hypoxic area to below 5000 km^2, as well as the load required to achieve that goal, has been in place for almost 15 years. Yet, almost no progress has been made (Sprague *et al.* 2011; Murphy *et al.* 2013). Similarly, water-quality improvement goals for the Chesapeake Bay have been in place for decades, but only recently has some measurable progress been achieved (USGS 2016). These and other cases highlight the need for persistent attention to sufficient implementation of proper management practices.

The health of Lake Erie was restored in the 1980s by addressing point source pollutants, and we are now faced with a similar problem from nonpoint source agricultural pollution that will require a very different solution. Fortunately, pathways to success exist but will require unparalleled collaboration and levels of implementation. Scientists, managers, and policy makers must develop sustainable and balanced solutions to meet this challenge.

Acknowledgements

This work was supported by the Erb Family Foundation grant 856, US National Science Foundation (NSF) grant 1313897, Joyce Foundation grant 15-36415, and NOAA grant NA13OAR4310142 to the University of Michigan; a Great Lakes Protection Fund grant and an International Plant Nutrition Institute grant 58-3604-4-005 to LimnoTech; an NSF grant GRT00022685 and an Ohio Sea Grant to Ohio State University; USDA agreement 59-3604-4-001 through USDA agreement 58-3604-4-005 to Heidelberg University; and USDA support for the ARS model. We thank J Arnold and M White for their thoughtful insights and advice throughout the project; representatives of the environmental and agricultural communities for helping shape the questions we asked of the models; and E Cisar, D Saad, T Stuntebeck, and T Bruulsema for their comments on this work. USGS was not involved in the development of the models used in this study. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the US Government.

References

Arnold J, Srinivasan MR, and Williams J. 1998. Large-area hydrologic modeling and assessment: part I. Model development. *J Am Water Resour As* **34**: 73–89.

Baker DB, Confesor R, Ewing DE, *et al.* 2014. Phosphorus loading to Lake Erie from the Maumee, Sandusky and Cuyahoga rivers: the importance of bioavailability. *J Great Lakes Res* **40**: 502–17.

Bierman VJ. 1980. A comparison of models developed for phosphorus management in the Great Lakes. In: Loehr RC, Martin CS, and Rast W (Eds). Phosphorus management strategies for lakes. Ann Arbor, MI: Ann Arbor Science Publishers.

Boomer KMB, Weller DE, Jordan TE, *et al.* 2013. Using multiple watershed models to predict water, nitrogen, and phosphorus discharges to the Patuxent Estuary. *J Am Water Resour As* **49**: 15–39.

Breuer L, Huisman JA, Willems P, *et al.* 2009. Assessing the impact of land use change on hydrology by ensemble modeling (LUCHEM). I: Model intercomparison with current land use. *Adv Water Resour* **32**: 129–46.

Bridgeman TB, Chaffin JD, and Filbrun JE. 2013. A novel method for tracking western Lake Erie *Microcystis* blooms, 2002–2011. *J Great Lakes Res* **39**: 83–89.

Bruulsema T, Lemunyon J, and Herz B. 2009. Know your fertilizer rights. *Crops Soils* **42**: 13–18.

Chaffin JD, Bridgeman TB, and Bade DL. 2013. Nitrogen constrains the growth of late summer cyanobacterial blooms in Lake Erie. *Adv Microbiol* **3**: 11.

Charlton MN, Milne JE, Booth WG, *et al.* 1993. Lake Erie offshore in 1990 – restoration and resilience in the central basin. *J Great Lakes Res* **19**: 291–309.

Diaz RJ and Rosenberg R. 2008. Spreading dead zones and consequences for marine ecosystems. *Science* **321**: 926–29.

Duan Q, Ajami NK, Gao X, and Sorooshian S. 2007. Multi-model ensemble hydrologic prediction using Bayesian model averaging. *Adv Water Resour* **30**: 1371–86.

Engel B, Storm D, White M, *et al.* 2007. A hydrologic/water quality model application protocol. *J Am Water Resour As* **43**: 1223–36.

Fales M, Dell R, Herbert ME, *et al.* 2016. Making the leap from science to implementation: strategic agricultural conservation in Michigan's Saginaw Bay watershed. *J Great Lakes Res* **42**: 1372–85.

GLWQA (Great Lakes Water Quality Agreement). 2012. The 2012 Great Lakes Water Quality Agreement. http://tinyurl.com/gt92hrh. Viewed 25 Feb 2016.

GLWQA (Great Lakes Water Quality Agreement). 2015. Recommended phosphorus loading targets for Lake Erie. Annex 4 Objectives and Targets Task Team Final Report to the Nutrients Annex Subcommittee. May 2015. http://tinyurl.com/pcvwt8g. Viewed 19 Jan 2017.

GLWQA (Great Lakes Water Quality Agreement). 2016. The United States and Canada adopt phosphorus load reduction targets to combat Lake Erie algal blooms. https://binational.net/2016/02/22/finalptargets-ciblesfinalesdep. Viewed 25 Feb 2016.

Han H, Allan DJ, and Bosch NS. 2012. Historical pattern of phosphorus loading to Lake Erie watersheds. *J Great Lakes Res* **38**: 289–98.

IJC (International Joint Commission). 2014. A balanced diet for Lake Erie: reducing phosphorus loadings and harmful algal blooms. Report of the Lake Erie Ecosystem Priority. Windsor, Canada: IJC.

Kalcic MM, Kirchhoff C, Bosch N, *et al.* 2016. Engaging stakeholders to define feasible and desirable agricultural conservation in western Lake Erie watersheds. *Environ Sci Technol* **50**: 8135–45.

Keitzer SC, Ludsin SA, Sowa SP, *et al.* 2016. Thinking outside the lake: can controls on nutrient inputs into Lake Erie benefit stream conservation in the watershed? *J Great Lakes Res* **42**: 1322–31.

Ludsin SA, Kershner MW, Blocksom KA, *et al.* 2001. Life after death in Lake Erie: nutrient controls drive fish species richness, rehabilitation. *Ecol Appl* **11**: 731–46.

Maccoux MJ, Dove A, Backus SM, and Dolan DM. 2016. Total and soluble reactive phosphorus loadings to Lake Erie: a detailed accounting by year, basin, country, and tributary. *J Great Lakes Res* **42**: 1151–65.

Michalak AM, Anderson EJ, Beletsky D, *et al.* 2013. Record-setting algal bloom in Lake Erie caused by agricultural and meteorological trends consistent with expected future conditions. *P Natl Acad Sci USA* **110**: 6448–52.

Moriasi DN, Arnold JG, Van Liew MW, *et al.* 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Trans ASABE* **50**: 885–900.

Muenich RL, Kalcic M, and 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**: 8146–54.

Murphy JC, Hirsch RM, and Sprague LA. 2013. Nitrate in the Mississippi River and its tributaries, 1980–2010 – an update: US Geological Survey Scientific Investigations Report 2013–5169. Reston, VA: USGS.

NRCS (Natural Resources Conservation Service). 2016. US effects of conservation practice adoption on cultivated cropland acres in Western Lake Erie Basin, 2003–06 and 2012. http://bit.ly/2iPwsmj. Viewed 19 Jan 2017.

Obenour D, Gronewold A, Stow CA, *et al.* 2014. Using a Bayesian hierarchical model to improve Lake Erie cyanobacteria bloom forecasts. *Water Resour Res* **50**: 7847–60.

Paerl HW and Paul VJ. 2012. Climate change: links to global expansion of harmful cyanobacteria. *Water Res* **46**: 1349–63.

Raftery AE, Gneiting T, Balabdaoui F, and Polakowski M. 2005. Using Bayesian model averaging to calibrate forecast ensembles. *Mon Weather Rev* **113**: 1155–74.

Scavia D, Allan JD, Arend KK, *et al.* 2014a. Assessing and addressing the re-eutrophication of Lake Erie: Central Basin hypoxia. *J Great Lakes Res* **40**: 226–46.

Scavia D, DePinto JV, Zhang H, *et al.* 2014b. Great Lakes Water Quality Agreement Nutrient Annex 4 Objectives and Targets Development Task Team Multi-Modeling Report – final. http://tinyurl.com/jynqcn3. Viewed 19 Jan 2017.

Scavia D, DePinto JV, and Bertani I. 2016a. A multi-model approach to evaluating target phosphorus loads for Lake Erie. *J Great Lakes Res* **42**: 1139–50.

Scavia D, Justic D, and Bierman VJ. 2004. Reducing hypoxia in the Gulf of Mexico: advice from three models. *Estuaries* **27**: 419–25.

Scavia D, Kalcic M, Muenich RL, *et al.* 2016b. Informing Lake Erie agriculture nutrient management via scenario evaluation. Ann Arbor, MI: University of Michigan. http://tinyurl.com/je56uel. Viewed 19 Jan 2017.

Seiller G, Anctil F, and Perrin C. 2012. Multimodel evaluation of twenty lumped hydrological models under contrasted climate conditions. *Hydrol Earth Syst Sci* **16**: 1171–89.

Sowa SP, Herbert M, Mysorekar S, *et al.* 2016. How much conservation is enough? Defining implementation goals for healthy fish communities in agricultural rivers. *J Great Lakes Res* **42**: 1302–21.

Sprague LA, Hirsch RM, and Aulenbach BT. 2011. Nitrate in the Mississippi River and its tributaries, 1980–2007: are we making progress? *Environ Sci Technol* **45**: 7209–16.

Stow CA, Roessler C, Borsuk ME, *et al.* 2003. Comparison of estuarine water quality models for Total Maximum Daily Load development in Neuse River Estuary. *J Water Resour Plan Manage* **129**: 307–14.

USGS (US Geological Survey). 2016. Summary of nitrogen, phosphorus, and suspended sediment loads and trends measured at the Chesapeake Bay nontidal network stations: water year 2014 update. http://tinyurl.com/jqk72w3. Viewed 29 Feb 2016.

Velazquez JA, Schmid J, Ricard S, *et al.* 2013. An ensemble approach to assess hydrological models' contribution to uncertainties in the analysis of climate change impact on water resources. *Hydrol Earth Syst Sci* **17**: 565–78.

Vollmer-Sanders C, Allman A, Busdeker D, *et al.* 2016. Building partnerships to scale up conservation: 4R Nutrient Stewardship Certification Program in the Lake Erie watershed. *J Great Lakes Res* **42**: 1395–402.

Weller DE, Benham B, Friedrichs M, *et al.* 2013. Multiple models for management in the Chesapeake Bay. STAC Publication Number 14-004. Edgewater, MD.

Yen H, Bailey RT, Arabi M, *et al.* 2014. The role of interior watershed processes in improving parameter estimation and performance of watershed models. *J Environ Qual* **43**: 1601–13.

Zhou Y, Michalak AM, Beletsky D, *et al.* 2015. Record-breaking Lake Erie hypoxia during 2012 drought. *Environ Sci Technol* **49**: 800–07.

Supporting Information

Additional, web-only material may be found in the online version of this article at http://onlinelibrary.wiley.com/doi/10.1002/fee.1472/suppinfo