WATER CENTER
UNIVERSITY OF MICHIGAN

APRIL 2016 UPDATE: See inside front cover for update information

Informing Lake Erie Agriculture Nutrient Management via Scenario Evaluation

UNIVERSITY OF MICHIGAN, ANN ARBOR

DONALD SCAVIA, MARGARET KALCIC, REBECCA LOGSDON MUENICH, NOEL ALOYSIUS, JEFFREY ARNOLD, CHELSIE BOLES, REMEGIO CONFESOR, JOSEPH DEPINTO, MARIE GILDOW, JAY MARTIN, JENNIFER READ, TODD REDDER, DALE ROBERTSON, SCOTT SOWA, YU-CHEN WANG, MICHAEL WHITE AND HAW YEN

Notes on this update:

- **None of the original results have been changed in this update.**
- **The numbering system and associated charts for the bundled scenarios have been changed to provide a more logical categorization of the results, and the associated text has been updated accordingly.**
- **The purpose of the ballpark phosphorus delivery pie charts has been clarified in the Introduction and Appendix A1.**
- **Description of how scenarios relate to baseline models has been included in Table 2.**
- **New tables have been added in Appendix A8 that describe impacts of the scenarios on watershed crop yields.**

Final Report

Informing Lake Erie Agriculture Nutrient Management via Scenario Evaluation

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

Compiled by the University of Michigan Water Center with funding from the Fred A. and Barbara M. Erb Family Foundation

[1] Water Center, Graham Sustainability Institute, University of Michigan, 625 E. Liberty, Ann Arbor MI 48103 USA

[2] Department of Food, Agricultural, and Biological Engineering, Ohio State University, Columbus, OH 43210, USA

[3] USDA, ARS, Grassland Soil and Water Research Laboratory, Temple, TX 76502, USA

[4] LimnoTech, 501 Avis Drive, Ann Arbor, MI 48108, USA

[5] National Center for Water Quality Research, Heidelberg University, Tiffin, OH 44883, USA

[6] Wisconsin Water Science Center, United States Geological Survey, 8505 Research Way Middleton, WI 53562, USA

[7] The Nature Conservancy, 101 E Grand River Ave., Lansing, MI 48906, USA

[8] Blackland Research & Extension Center, Texas A&M Agrilife Research, Temple, TX 76502, USA

Introduction

Harmful algal blooms (HABs) have been increasing in extent and intensity in the western basin of Lake Erie. The cyanobacteria *Microcystis* produces toxins that pose serious threats to animal and human health, resulting in beach closures and impaired water supplies, and have even forced a "do not drink" advisory for the City of Toledo water system for several days in the summer of 2014. The main driver of Lake Erie HABs is elevated phosphorus loading from watersheds draining to the western basin, particularly from the Maumee River watershed (Obenour et al. 2014). Through the 2012 Great Lakes Water Quality Agreement (GLWQA), the U.S. and Canadian governments agreed to revise Lake Erie phosphorus loading targets to decrease HAB severity below levels representing a hazard to ecosystem and human health. **New targets limit March-July loadings from the Maumee River to 186 metric tonnes of dissolved reactive phosphorus (DRP) and 860 metric tonnes of total phosphorus (TP) – a 40% reduction from 2008 loads** (GLWQA 2016).

The Great Lakes region must now determine what policy options are most effective and feasible for meeting those targets. While all sources are important, our focus is on agriculture because it overwhelms other sources. In a conservative ballpark estimate we found that 85% of the Maumee River's load to Lake Erie comes from farm fertilizers and manures, even though this is only 10% of farmland fertilizer applications (Figure 1). Load targets will not be met without reductions from agriculture.

Therefore, the overall goal of this study was to identify potential options for agricultural management to reduce phosphorus loads and lessen future HABs in Lake Erie. We applied multiple watershed models to test the ability of a series of land management scenarios, developed in consultation with agricultural and environmental stakeholders, to reach the proposed targets.

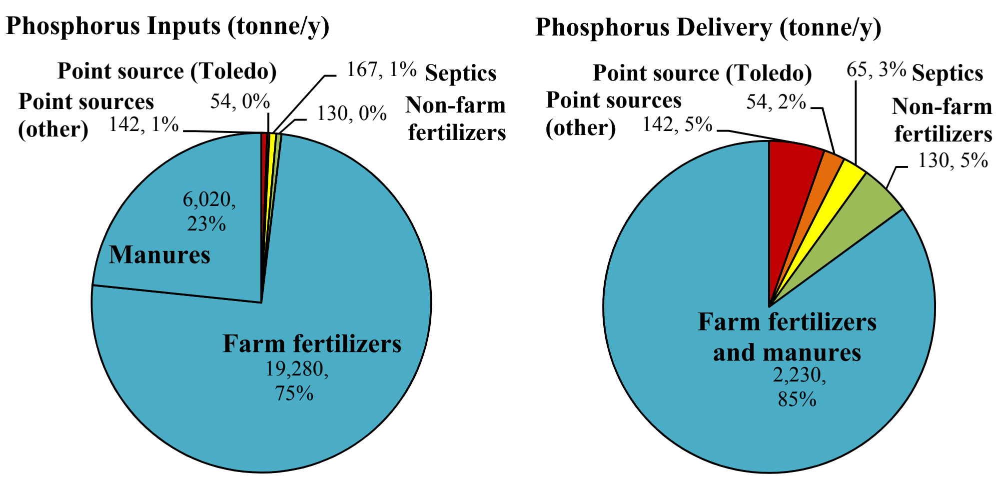


Figure 1: Maumee inputs and delivery of P to Lake Erie from major sources (Appendix A1). Estimated delivery from farm fertilizers and manures (2,230 t/y) is 10% of applied (25,300 t/y). This delivery was estimated conservatively with respect to agriculture by subtracting the known inputs of point sources, failing septic systems, and non-farm fertilizers (assuming 100 percent delivery to the lake) from the average Maumee River load 2005-2014. The delivered load from farm fertilizers and manures includes legacy sources in soils and streams. This estimate is illustrative, and these data were not used to drive the watershed models or any other results in this report.

Approach

Use of Multiple Watershed Models

We used multiple models to increase confidence in the results and provide decision-makers with a range of expected water quality outcomes. Each model has strengths and weaknesses, and many modeling decisions are subjective. Multi-model and ensemble modeling approaches have been applied in other fields (e.g. lake and estuary modeling: Weller et al. 2013; Scavia et al. 2004; Stow et al. 2003; IJC 1998; Bierman 1980; climate modeling: Tebaldi and Knutti 2007; and wildfire modeling: Yue et al. 2013), and a multi-model approach was used to support development of the new target phosphorus loads for the GLWQA (Scavia et al. in review; Scavia and DePinto 2015). Although ensemble modeling has been frequently applied to evaluate and compare hydrological predictions (Velazquez et al. 2013; Seiller et al. 2012; Breuer et al. 2009), very few studies have applied ensemble modeling for watershed water quality (Boomer et al. 2013), and none have applied used it to evaluate policy-relevant land management scenarios.

Because the Maumee watershed is critical to Lake Erie, we took advantage of several modeling groups that had previously developed and calibrated watershed models capable of testing agricultural management scenarios. The models include the Soil and Water Assessment Tool (SWAT) developed by five different modeling groups from Heidelberg University (HU) (Confesor et al. in prep), LimnoTech (LT) (Boles et al. in prep), Ohio State University (OSU) (Gildow et al. in review; Culbertson et al. in review; Gebremariam et al. 2014), Blackland Research & Extension Center, Texas A&M University (TAMU) (Keitzer et al. in review), and the University of Michigan (UM) (Kalcic et al. in review; Muenich et al. in review). In addition, the SPAtially Referenced Regressions On Watershed attributes (SPARROW) model developed by the U.S. Geological Survey (USGS) for the US-side of the Great Lakes region (Robertson and Saad 2011) was re-scaled to observed data from the Maumee River at Waterville, Ohio and included in the analysis.

While five of the six teams used SWAT, these are in fact different models because of the many independent critical decisions made about spatial discretization, input data sources, subroutines to use, land management operations, model parameterization, and calibration approaches (see Table 1). While there may be a temptation to select one model based on "superior performance," there are many ways to evaluate performance (e.g., graphical and statistical methods and ensuring field-level nutrient export, soil nutrient content, and crop yields are within observed ranges) and thus there is no unique measure of performance. Instead, we chose to use multiple models because the true accuracy of the models in representing the baseline condition is not uniquely quantifiable and each model gives a reasonable representation of the real world. When a range of models all project similar results, our confidence in those results increases significantly.

Description of the Watershed Modeling Tools

SWAT and SPARROW represent two different types of modeling approaches. SWAT is primarily a process-based model that represents watershed processes and interactions with physical and chemical equations. SPARROW is a hybrid mechanistic-statistical model, with empirically-based coefficients used to describe relationships between observed properties, such as measured runoff and landscape conditions. Each is described briefly below.

Table 1: Agricultural management scenarios were run in 5 separately-configured SWAT models. While all models were developed using the same base SWAT framework, they are each distinct in many ways, from initial model setup and activated model subroutines to assumptions about farmland management and model calibration. The main themes in model differences are shown below with examples of differences among models, explanations for the type of uncertainty from these differences, and a description of how models were homogenized for the Baseline. A model-specific list of differences among SWAT models can be found in Appendix A2. *For further information on farm management assumptions contact coauthors representing each of the modeling teams and associated citations (Confesor et al. in prep; Boles et al. in prep; Gildow et al. in review; Culbertson et al. in review; Gebremariam et al. 2014; Keitzer et al. in review; Kalcic et al. in review; Muenich et al. in review; Robertson and Saad 2011).

Aspect of modeling in order of development	**Potential differences among models in this study**	**Further details**
Spatial discretization & resolution *in initial model setup through ArcGIS interface*	• Size of sub-watersheds as dictated by stream threshold • Definition of HRU slope classes • Lumping of HRUs	Initial model set-up is determined based on the goals of the project, and once completed is difficult to change. These model differences were retained in the Baseline models. This source of uncertainty is referred to as structural uncertainty.
Model/Sub-model algorithms *chosen within SWAT*	• Model release version and source code updates • Tile drainage routine • In-stream processing • Evaporation method • Water table method • Runoff method • Carbon model • Soil phosphorus model	SWAT is a compilation of multiple sub-models, and the user can choose which sub-models to use. The algorithms used in the model introduce structural uncertainty.
Model inputs *including data sources, spatial resolution, and preprocessing*	• Land use data: NLCD vs. NASS CDL • Point source data: None included vs. included based on emissions caps vs. based on measured data • Weather data	Model inputs are also chosen early in the modeling process. In this study we chose to control for some of these input differences by homogenizing point sources and climate forcing across Baseline models. These choices introduce input and measurement uncertainty.
Land management operations *include a host of assumptions based on disparate sources**	• Spatial distribution/ heterogeneity of operations • Timing of operations • Crop rotations • Fertilizer applications • Manure applications • Inclusion of existing conservation practices	Assumptions made about cropland management operations are critically important for realistically simulating current agricultural practices in the watershed, many of which are difficult to determine using publicly available datasets. Cropland management differences were retained in the Baseline models. This is a form of input uncertainty, and addressing this was a primary goal of the study.
Model parameterization *in choosing realistic parameter values to calibrate a model*	• Parameters changed in calibration • Bounds on parameter values • Methods for assessing model performance during calibration	Modelers changed different sets of parameters to calibrate their models, and the final parameter values span a wide range. Multiple parameter sets can achieve a reasonably calibrated model, which leads to parameter uncertainty. These differences were retained in the Baseline models.
Measured data *for calibration*	• Extent of water quality calibration • Extent of hydrology considered at upstream monitoring stations • Method to fill in or ignore missing data	Measured data provides a reality check against which we assess how well our models perform. It is easy to forget that measured data are only a snapshot of true events and there can be considerable uncertainty in them.

The SWAT Model - SWAT is a semi-distributed, process-based, watershed-scale, hydrological model that uses inputs of soils, slope, land-use, land management information, and climate variables (precipitation, temperature, etc.) to estimate hydrology, water quality, and plant growth (Arnold et al. 1998). SWAT performs daily calculations and provides outputs at many spatial scales (field, river, sub-watershed) that can be aggregated to many temporal scales (daily to decadal). The smallest spatial scale is the hydrologic response unit (HRU), which is a combination of unique soils, slopes, and land-uses within a sub-watershed. HRU outputs are added together and routed through the reaches (streams and rivers) to the watershed outlet.

Within the overall architecture of SWAT, there are many algorithms for calculating storage and flux of water and nutrients, and many sub-models from which to choose (Table 1). These sub-models include different approaches for estimating the water balance (evapotranspiration, water table depth, tile drainage, and runoff) as well as soil and nutrient transformations (in-soil carbon and phosphorus models, as well as in-stream nutrient processing). A single difference in a sub-model can influence hydrologic pathways and the fate of nutrients throughout the watershed.

Many cropland management options are possible in the model, making it particularly well suited to applications in agricultural watersheds (Gebremariam et al. 2014; Douglas-Mankin et al. 2010; Van Liew et al. 2007), and enabling scenario testing of those options. Thus, many independent decisions are made about cropland management options to include in the models, and these decisions introduce variability among model applications. Modelers also chose from different data sources about cropland management, including: fertilizer application rates, timing, type, and method; type and timing of tillage operations; crop rotations; existing conservation practices; and the spatial distribution of these practices across a watershed (Table 1).

SWAT is typically calibrated to monitoring data at one or more gaging station(s) in the watershed. Most models are calibrated to streamflow, and some are calibrated to concentrations or loads of nutrients and sediments when measured water quality data are available. Commonly used measures of goodness-of-fit include the coefficient of determination (R^2), the Nash-Sutcliffe efficiency (NSE), and percent bias (Engel et al. 2007; Moriasi et al. 2007). Additional calibration efforts include ensuring model processes are producing realistic results, including confirming that crop yields and the partitioning of streamflow sources from surface flow, tile drainage flow, and base flow recharge are within observed ranges (Wellen et al. 2015; Yen et al. 2014a; Yen et al. 2014b). During calibration, model parameters that drive hydrology and water quality are changed iteratively, and because SWAT has many such parameters, unique combinations may be able to produce the same quality of calibration. It is not possible to know with certainty which combinations of parameter values are the most correct, and this is called parameter uncertainty (Table 1).

The SPARROW Model - SPARROW is a watershed model that uses a mass-balance approach to estimate the non-conservative transport and transformation (i.e., losses) of nutrients under long-term steady-state conditions in relation to statistically significant landscape properties, such as climate, soils, and artificial drainage (Robertson et al. 2009; Schwarz et al. 2006). SPARROW is a spatially explicit model that estimates nutrient loading from a series of hydrologically linked catchments. SPARROW models simulate long-term mean-annual nutrient transport given nutrient inputs similar to a base year (for use in this study, it was calibrated for inputs similar to 2002). Appendix A3 provides details on the SPARROW model input.

The USGS's Midwest SPARROW models were developed to describe the delivery of phosphorus (P) and nitrogen (N) throughout the Upper Midwest, including all U.S. drainages to the Great Lakes (Robertson and Saad 2011). The main purpose was to: 1) determine P and N loads to each Great Lake (from the U.S. part of their basins); 2) determine the total P and N load from each tributary draining more than 150 km^2 to each Great Lake; 3) rank the individual tributaries to each lake based on their relative loading and yields; 4) determine the relative importance of each P and N source; and 5) determine which environmental factors significantly affect the delivery of P and N from the land to the streams in the Upper Midwest.

In general, SPARROW models are calibrated by minimizing the error between observed and estimated long-term average annual loads in natural log units using nonlinear regression. Individual source variables are typically included in the model only if they are statically significant ($p < 0.05$) in explaining variation in P and N loads. In instances where specific source variables known to be important are not significant, they are usually combined with other similar variables to create composite variables through a series of calibration runs until an acceptable level of model fit is achieved, as measured by root mean square error (RMSE), coefficient of determination (R^2), variance inflation factors (VIF), model-estimated coefficients, and spatial distributions of residual errors.

Establishing a common baseline for validation and scenario comparisons

All SWAT models were previously calibrated and validated (for details see Appendix Table A2.1), and in this work we verified that model performance was still acceptable for a common baseline time (2005-2014) period using the same tests as described above for calibration (R^2, Nash-Sutcliffe Efficiency, percent bias, graphical time-series comparisons, and other calibration checks). To control for some input uncertainty and to eliminate some of the variance among models, all SWAT models received the same precipitation, temperature, and point source data (Appendices A4 and A5) for model validation and as a baseline against which the scenarios were compared. The 2005-2014 baseline time period was used for model validation and scenarios because this corresponds to the elevated HAB issue in Lake Erie. The measured data for validation was taken from USGS and Heidelberg datasets for Waterville, Ohio, and daily loads were estimated from daily concentrations and flows (Appendix A6). Monthly loads were summed from daily estimates. Missing values were replaced using a method specifically designed for this dataset (Obenour et al. 2014), and months with more than two weeks of missing days were excluded for validation. Output from the SPARROW model was modified to reflect common point source data (2011), and non-point source delivery rates were rescaled to reproduce the average annual measured TP loading (detrended to 2011) at the Waterville, Ohio gaging station. The full procedure used to modify SPARROW output to represent recent conditions is summarized in Appendix A3.

Developing Land Management Scenarios

We developed potential scenarios for agricultural land management through in-depth conversations among modelers and experts from the agricultural and environmental communities (Table 2). Modelers provided information about what scenarios were feasible and best able to be tested with the models. The environmental and agriculture experts provided insights into practical implementations and policy feasibility.

Table 2: Description of the bundled scenarios. Practices were applied in the specified percentage of cropland, with baseline practices used in the remaining croplands. For additional modeling details see Appendix Table A8.2.

No.	Name	Description of scenarios	Relationship to baseline
1	**No Point Source Discharges**	All PS discharges were removed (i.e., set to zero).	Baseline models had point sources, which were removed in this scenario.
2a-c	**Cropland conversion to grassland** at 10% (2a), 25% (2b), and 50% (2c) targeted adoption	In these three scenarios designed to test how much land would need to be removed from production if farms adopted no additional conservation practices, 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.	Baseline models considered all cropland to be cultivated. In this scenario a percentage of that cropland was converted to switchgrass.
3	**In-field practices** at 25% random adoption	The following practices were applied together on a random 25% of row cropland: 50% reduction in P fertilizer application, fall timing of P applications, subsurface placement of P fertilizers, and a cereal rye cover crop.	Baseline models included a wide range of assumptions about P fertilizer and manure application rates (from low to high), timing (from fall to spring), and placement (from broadcast and incorporated to primarily subsurface applied). Therefore, in some models, nutrient management scenarios diverged more from the baseline than others, resulting in a range of predicted water quality benefits. As with all scenarios, cropland untouched by a scenario retained baseline practices. Baseline models did not include winter cover crops (other from winter wheat), nor did they include existing buffer strips. Therefore, results from those scenarios call for additional percentage of cover crops and buffer strips.
4	**Nutrient management** at 25% random adoption	The following practices were applied to a randomly selected 25% of row crop acreage: a 50% reduction in P fertilizer application, fall timing of P applications, and subsurface placement of P into the soil.	
5	**Nutrient management** at 100% adoption	The following practices were applied to 100% of row crop fields: a 50% reduction in P fertilizer application, fall timing of P applications, and subsurface placement of P into the soil.	
6	**Commonly recommended practices** at 100% random adoption	The following 4 practices were each applied to separate 25% of the crop acres: a 50% reduction in P fertilizer application, subsurface application of P fertilizers, continuous no-tillage, and medium-quality buffer strips.	
7	**Continuous no-tillage and subsurface placement of P fertilizer** at 50% random adoption	A combination of continuous no-tillage and subsurface application of P fertilizers were applied together on a randomly selected 50% of row crop acres.	
8	**Series of practices** at 50% targeted adoption	The following practices were targeted to the 50% of row cropland with the highest TP loss in the watershed: subsurface application of P fertilizers, cereal rye cover crop in the winters without wheat, and application of medium-quality buffer strips.	
9	**Series of practices** at 50% random adoption	The following practices were applied to a random 50% of row cropland: subsurface application of P fertilizers, cereal rye cover crop in the winters without wheat, and application of medium-quality buffer strips.	
10	**Diversified rotation** at 50% random adoption	An alternative corn-soybean-wheat rotation with a cereal rye cover crop all winters without wheat was applied over a randomly chosen 50% of row cropland.	Baseline models had a rotation containing wheat, which in this scenario was applied in rotation with rye in 50% of farm fields.
11	**Wetlands and buffer strips** at 25% targeted adoption	Wetlands treating half of overland flow in a sub-watershed were targeted to 25% of sub-watersheds with the greatest TP loading rates and medium-quality buffer strips were targeted to 25% of row cropland with greatest TP loss rates.	Most baseline models had no wetlands or buffers; those present remained or, if targeted, were replaced.

Models were first run to test single-practice scenarios (Appendix A7) to explore the bounds of what might be possible and to provide a first-look comparison among models. Based on these results, "bundles" (Table 2) of single practices were tested at various rates of implementation. Modeling details associated with each bundle are provided in Appendix A8. In general, the bundles ranged from implementation on 25% to 100% of cropland, practices were implemented either randomly or targeted to locations that each model simulated to be physically vulnerable to TP loss or having lower crop yields, and bundles were developed in the context of specific policy questions. Results of bundles simulated with SWAT were reported in comparison to the baseline scenarios from each model.

Results

Validation of Baseline Models - All models performed well in the 2005-2014 validation period (Table 3). The TAMU and HU models were recalibrated somewhat to improve simulation of DRP, but the other models performed well without additional calibration. The models reproduced flow and P loading from the Maumee River to Lake Erie. SWAT models differed in predicting inter-annual March-July phosphorus loading (Figure 2); however, the multi-model average is close to observations, particularly for TP. Many models also simulated DRP well; however, the overall average slightly over-estimates DRP delivery for low loading levels. While all model-simulated streamflow and P loadings were within accepted norms for this type of study (Engel et al. 2007; Moriasi et al. 2007), we chose to remove any remaining biases when comparing scenarios to inform options to reach the new loading targets. Therefore, to calculate each model's responses to the scenarios, we multiplied the percent change between scenario and baseline loadings for each model by the average observed 2005-2014 loadings at Waterville, Ohio.

Table 3: SWAT validation for monthly flow and phosphorus loading at Waterville, Ohio, near the Maumee River outlet to Lake Erie. Percent bias (PBIAS) is a measure of how much a model overestimates or underestimates flow and phosphorus loading over the entire period (values closer to 0 indicate better agreement). Nash-Sutcliffe Efficiency (NSE) and the coefficient of correlation (R^2) indicate how closely the monthly flows and loads correspond to measured data (values closer to 1 indicate better agreement). The detrended annual TP loads used in the SPARROW model were 4% less than the average of the 2005-2014 period (not shown).

	Measure of model fit	**Criterion for excellent fit**	Heidelberg University **(HU)**	LimnoTech **(LT)**	Ohio State University **(OSU)**	Texas A&M University **(TAMU)**	University of Michigan **(UM)**	**Model average**
Flow	PBIAS	+/- 10%	-7%	10%	10%	11%	6%	6%
	NSE	> 0.5	0.82	0.90	0.91	0.86	0.89	0.88
	R^2	> 0.6	0.86	0.91	0.93	0.88	0.91	0.90
TP	PBIAS	+/- 25%	37%	-6 %	-7%	-22%	7%	2%
	NSE	> 0.4	0.64	0.82	0.73	0.56	0.70	0.69
	R^2	> 0.5	0.74	0.82	0.75	0.71	0.70	0.75
DRP	PBIAS	+/- 25%	81%	1 %	16%	-13%	-13%	14%
	NSE	> 0.4	-0.02	0.71	0.51	0.52	0.46	0.44
	R^2	> 0.5	0.55	0.71	0.54	0.70	0.51	0.60

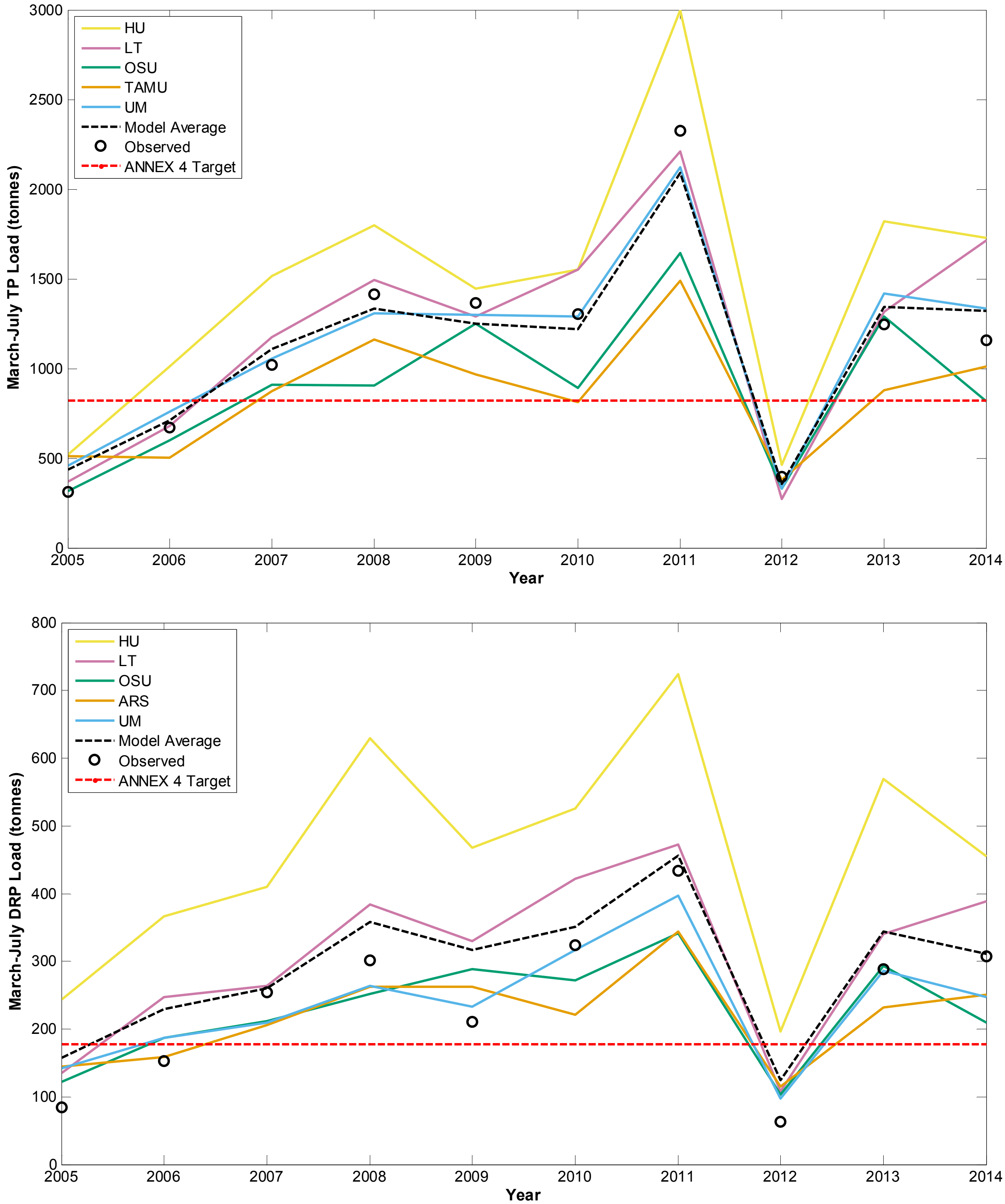


Figure 2: Inter-annual performance of SWAT models in predicting March-July TP (top) and DRP (bottom) loads.

Exploring Potential Phosphorus Hotspots - While models had similar predictions at the watershed outlet, their varying assumptions about land management across the watershed resulted in different estimates of what parts of the watershed contribute the most P. Delivered yields to Lake Erie from sub-watersheds was compared among the baseline SWAT and SPARROW models to estimate locations that, if untreated by conservation practices, would be most vulnerable to contributing P to Lake Erie. Delivery to the Lake was calculated by partitioning the load at the outlet to upstream sub-watersheds based on their relative loadings (see Appendix A9 for details). Vulnerable locations, or potential hotspots, were defined as sub-watersheds within the highest 20th percentile of delivered P yield to the lake (kg/km^2). Vulnerability maps were prepared by summing the number of models that agreed that a particular sub-watershed is a potential hotspot (Figure 3). Agreement among models was greater for TP than DRP because the models use different assumptions about the sensitive partitioning of DRP between surface and sub-surface flows, as well as the location and characteristics of tile drains and crop rotations. Additional baseline model results can be found in Appendix A10.

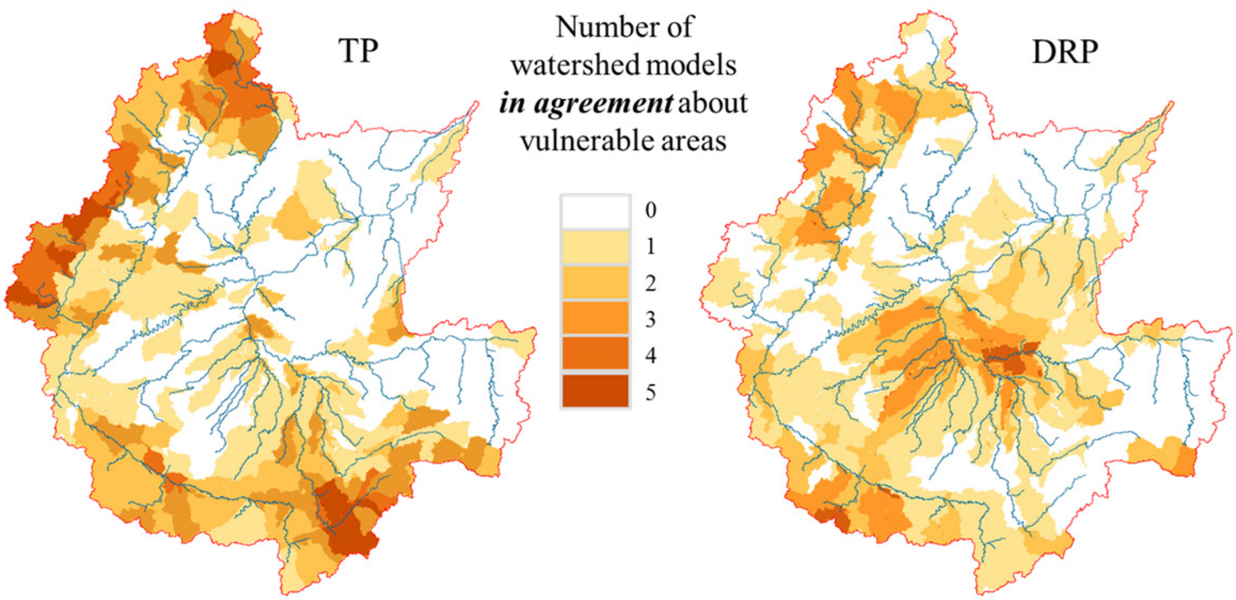


Figure 3: Vulnerable areas or potential "hotspots" identified by Baseline SPARROW and SWAT models. The scale is 0-to-5 as the 6 models never fully agreed on TP hotspots. The SPARROW model was not used for DRP.

Land Management Scenarios

All bundled land management scenarios reduced both TP and DRP delivery to the lake, with larger reductions from greater implementation and targeting (Figure 4). However, not all scenarios were able to meet the targets of 186 metric tonnes of DRP and 860 metric tonnes of TP delivered from the Maumee River in March-July. Here we report the average and standard deviation among SWAT models (results for the individual models are in Appendix A8) because SPARROW was not designed to test these bundled scenarios. Findings in relation to specific policy questions are provided in Table 4.

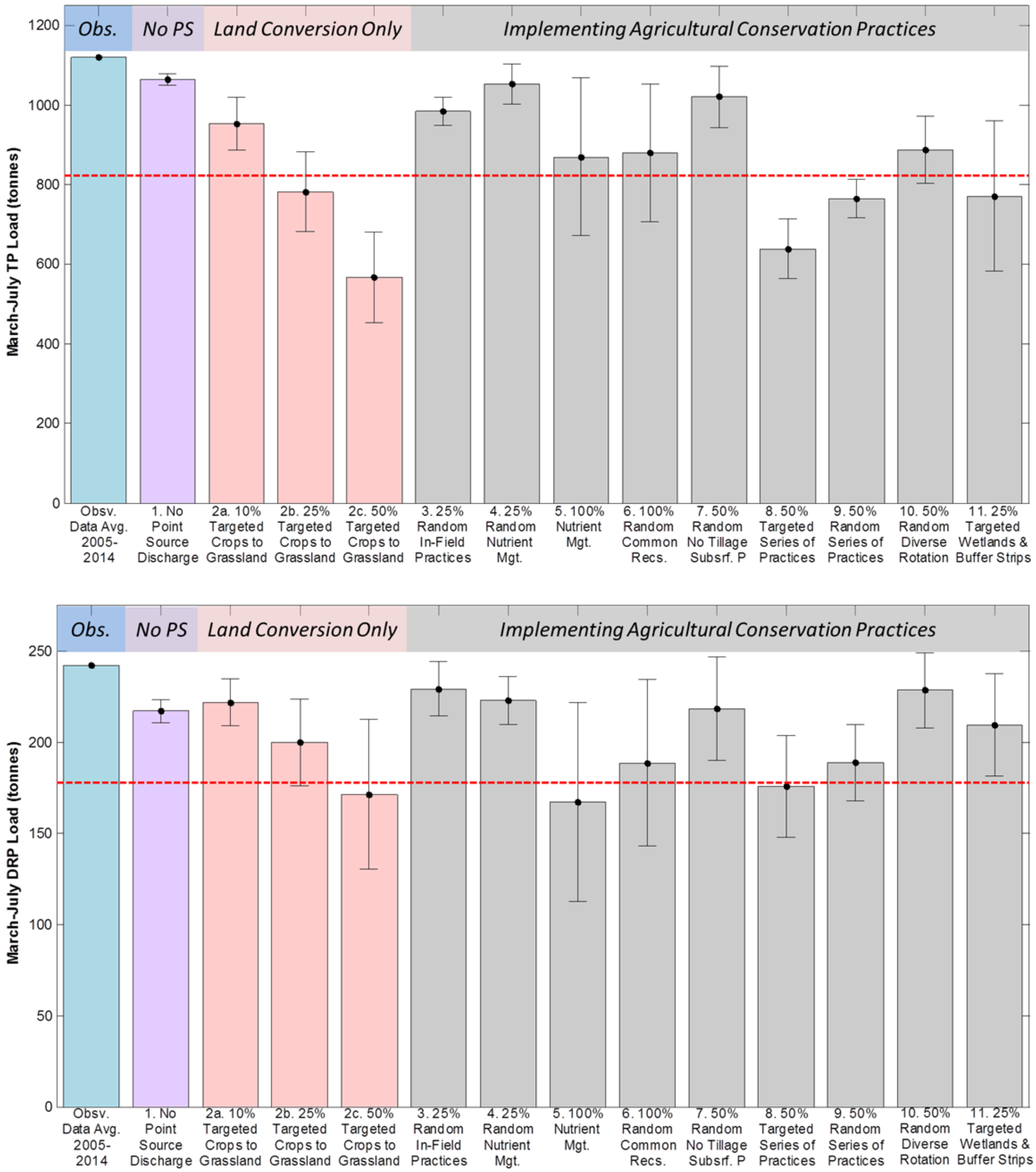


Figure 4: Average and standard deviation of the five SWAT models' March-July TP (top) and DRP (bottom) loads during the 2005-2014 modeling time period. The average observed March-July loads from 2005-2014 are shown in the blue bars, the result for removing all point source discharges in the watershed is shown in the purple bars, and the GLWQA target loads (area-weighted to Waterville, OH gage station) are shown by the red dashed lines. Pink bars 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. Gray bars show the effect of implementing more agricultural conservation. Corresponding values are provided in Table A8.5.

Extreme scenario 1 that eliminates all point source discharges reduced the March-July TP and DRP loads by only 5% and 10%, respectively, illustrating the significance of agricultural sources. The land conversion scenarios (2a-c) are rather extreme scenarios that are unlikely to be implemented. They were included to illustrate how much 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 (Appendix Tables A8.3 and A8.4). Overall, the most promising scenarios included widespread use of nutrient management practices, especially subsurface application of P fertilizers, which was the most helpful single practice for DRP (Appendix Figure A7.4), and installation of buffer strips.

Discussion

Care must be taken in interpreting these results because some portions of our scenarios may already be implemented to some degree within the watershed. However, because of privacy issues, we were not able to determine the extent or location of buffer strips, winter cover crops aside from wheat, and wetlands. For these practices, the best interpretation of our results is that they identify the need for ***additional implementation.*** For example, to achieve a result like scenario 9, an additional 50% of cereal rye and buffer strips are required. Current estimates are that 8% and 35% of farms currently apply these practices, respectively, in this watershed (Wilson et al. 2013). Other existing practices such as timing of P applications, subsurface placement of P, continuous no-tillage, winter wheat grown in rotation, and fertilizer application rates are included to some extent in the baseline models. The best interpretation of those results, as well as for land conversion to switchgrass, is that they identify the required ***total level of implementation.*** Appendix Table A8.1 provides details on the extent of implementation for each practice to aid in scenario result interpretation.

Our results suggest that there are pathways to achieve the new target loads for Lake Erie. However, all of the successful pathways require significant levels of implementation of both common and less common practices. For example, three scenarios that appear to be able to reach the TP goal (Figure 4) simulated both targeted (scenario 8) and random (scenario 9) treatment of 50% of croplands with a combination of nutrient management and in-field (cover crops) and edge of field practices (buffer strips) or a combination of wetland and buffer strip installations on 25% of cropland or subbasins, respectively (scenario 11). These scenarios also highlight the importance of placing practices in areas where they are needed most. While identifying these specific locations was beyond the scope of this work, it can be done in consultation with conservationists and producers that have intimate knowledge of farm landscapes.

Scenarios 8 and 5 achieved the DRP target loads (Figure 4). Scenario 5, which simulated implementation of nutrient management practices on 100% of the cropland acres, supports the importance of the right rate and right placement of P applications promoted by the Western Basin 4R Nutrient Stewardship Certification Program that was launched in 2014 which certified nutrient management plans on 26% of the cropland in the basin in just two years (Vollmer-Sanders et al. in press). Scenario 5 also produced TP reductions near the 40% goal.

Table 4: Summary of the project findings according to the policy questions they were intended to address.

No.	Name	Policy question	Project Findings
1	**No Point Source Discharges**	Can phosphorus targets be reached by point source management alone?	Removing point sources entirely from the watershed reduced phosphorus loading, but did not achieve targets.
2a-c	**Cropland conversion to grassland** at 10% (2a), 25% (2b), and 50% (2c) targeted adoption	If agricultural management is unchanged, how much row cropland would need to be converted to grassland to reach the targets?	In this dose-response approach, we found that TP targets could be achieved with nearly 25% conversion of cropland to grassland, and DRP targets were met with closer to 50% conversion. The difficulty reducing DRP loadings may be a result of legacy P stored in soils within the Maumee River watershed.
3	**In-field practices** at 25% random adoption	What can be achieved at 25% application of in-field practices?	While in-field practices did serve to reduce both TP and DRP losses, random implementation on only 25% of croplands was not enough to achieve either the TP or DRP targets.
4	**Nutrient management** at 25% random adoption	What level of nutrient management will be sufficient to reach phosphorus targets?	Nutrient management at 25% implementation is not enough to achieve TP or DRP load targets.
5	**Nutrient management** at 100% adoption	Can nutrient management alone achieve targets?	On average, nutrient management alone has the potential to achieve DRP targets, but not TP targets.
6	**Commonly recommended practices** at 100% random adoption	What extent of adoption of commonly recommended practices will be needed to achieve the targets?	While 100% adoption of at least one commonly recommended conservation practice helped move average loads closer to target goals, adoption of multiple practices per farm field may be required to achieve the targets.
7	**Continuous no-tillage and subsurface placement of P fertilizer** at 50% random adoption	Is no-tillage effective provided P is applied below the soil surface?	Implementing subsurface application of P fertilizers in a no-tillage system can help reduce P losses; however, when implemented on 50% of cropland, this combination of practices is not sufficient to achieve load targets.
8	**Series of practices** at 50% targeted adoption	What extent of targeted in-field and edge-of-field practices reaches the targets?	Results showed that a series of in-field and edge-of-field practices on the same crop fields could achieve the TP load target with random application at 50% adoption and well exceeded the target load with targeted placement of the practices on high P exporting croplands. Targeted implementation was required to achieve the DRP target load. These results indicate the value of targeting conservation practices to lands with the highest P losses.
9	**Series of practices** at 50% random adoption	What if in-field and edge-of-field practices were applied at random?	
10	**Diversified rotation** at 50% random adoption	What is the impact of returning to winter wheat and winter cover crops?	The results of the diversified rotations are less conclusive as some of the models had Baseline wheat rotations where the wheat was double-cropped with soybean in the same year. On average, the models showed marked reductions in TP loads and some improvement in DRP loads with the diversified rotation.
11	**Wetlands and buffer strips** at 25% targeted adoption	How much P reduction can be achieved through structural practices?	Wetlands targeted to 25% of high P loading sub-watersheds and buffer strips targeted to 25% of high P exporting cropland could achieve TP loading targets on average, but not DRP. This is partially due to the fact that much of DRP exits cropland via subsurface drains which are not intercepted by buffer strips.

While not all potential practices or combinations of practices were simulated in this work, it is clear that reaching the new target loads is a daunting challenge and will require large changes in management and much greater investment of resources to achieve the required levels of implementation, particularly for the less commonly applied practices. These results are consistent with 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. in press) and the Western Basin (NRCS 2016; Keitzer et al. in review; Muenich et al. in review; Kalcic et al. in review). Results across these studies clearly show that funding levels within the conservation provision of the current Farm Bill are currently alone insufficient to address these problems. What is needed now is for key local, state, and federal management agencies and the public and private sector to come together and use the information from these studies to help set shared implementation goals and to demand innovation and honest assessments of existing and potentially new programs, policies, and partnerships that will be able to achieve these stretch goals. Fortunately, there are some innovative efforts like water funds, pay-for-performance, and public-private partnerships underway within the Western Basin of Lake Erie and other parts of the Great Lakes that are moving us in this direction (Fales et al. in press). NRCS's recent three-year $41M investment to target, expand, and accelerate conservation practices in the Western Basin is a substantial 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 see measureable improvements in water quality.

Historically, agricultural conservation efforts have sought to reduce soil erosion, and more recently nitrogen export from farmland through voluntary implementation of practices. At this time, it is not clear if current programs have sufficient funding or policies in place that enable targeting of the best practices in the right places to support implementation at the necessary scale to reduce *phosphorus* export. The difficulty in reaching load reduction targets has precedent from other regions. For example, while the goal of reducing the Gulf of Mexico hypoxic area to below 5,000 km^2, as well as the load reduction required to achieve that goal, have been in place for 15 years, almost no progress has been made (Sprague et al. 2011, Murphy et al. 2013) under current programs. Similarly, water quality improvement goals for the Chesapeake Bay were in place for decades before some limited progress was made (USGS 2016), but this required the states to partner with the USEPA to go beyond the current Farm Bill and similar conservation programs and implement stronger nutrient management to comply with the Clean Water Act.

Finally, our results also indicate that even with extensive implementation across row cropland, the scenarios that meet the target on average may not meet it in every year, especially in years with above-average precipitation or extensive snowmelt. There may also be time lags between the timing of practice adoption and the loading of legacy P sources. While additional research is needed to more fully quantify the influence of projected climate and rates of practice adoption, most climate models project changes in precipitation for this region. These changes in precipitation may make progress more challenging in the future.

Acknowledgements

This project was supported by an Fred A. and Barbara M. Erb Family Foundation grant (Grant # 856), a NSF Watershed Sustainability Climate grant (Grant # 1313897), a Joyce Foundation grant (Grant # 15-36415), and a NOAA Coastal and Oceanic Climate Applications grant (Grant # NA13OAR4310142) to the University of Michigan; an NSF Coupled Human and Natural Systems grant (GRT00022685) and a NOAA/Ohio Sea Grant to Ohio State University; U.S. Army Corps of Engineers – Buffalo District task orders, a Great Lakes Protection Fund grant (Project No. 936.01), and an International Plant Nutrition Institute grant through the 4R Nutrient Stewardship Research Fund (Agreement # 58-3604-4-005) to LimnoTech; a 4R-Research Project sub-award from USDA Agricultural Research Service (Agreement # 59-3604-4-001) with the prime grant from the International Plant Nutrition Institute made to the USDA-ARS (Agreement # 58-3604-4-005) to Heidelberg University; and USDA support to the Blackand Research & Extension Center, Texas A&M University. The authors would also like to thank the environmental and agricultural communities for helping to shape the questions we asked of the models, and Elizabeth Cisar of the Joyce Foundation, David Saad of USGS, and Tom Bruulsema of IPNI for reviewing this report.

References

Arnold JG, Srinivasan R, Muttiah RS, Williams JR. 1998. Large area hydrologic modeling and assessment part 1: Model development. Journal of the American Water Resources Association, 34(1): 73-89. doi: 10.1111/j.1752-1688.1998.tb05961.x.

Bierman VJJ. 1980. “A Comparison of Models Developed for Phosphorus Management in the Great Lakes” *in Phosphorus Management Strategies for Lakes*, RC Loehr, CS Martin and W Rast, Eds. Ann Arbor Science Publishers, Ann Arbor, MI. pp. 235-255.

Boles C, Redder T, DePinto JV. Use of a Calibrated SWAT Model to Support Best Management Practice (BMP) Evaluations in the Maumee River Watershed. In preparation.

Boomer KMB, Weller DE, Jordan TE, Linker L, Liu Z-J, Reilly J, Shenk G, Voinov AA. 2013. Using multiple watershed models to predict water, nitrogen, and phosphorus discharges to the Patuxent Estuary. Journal of the American Water Resources Association, 49(1): 15-39.

Breuer L, Huisman JA, Willems P, Bormann H, Bronstert A, Croke BFW, Frede H-G, Gräff T, Hubrechts L, Jakeman AJ, Kite G, Lanini J, Leavesley G, Lettenmaier DP, Lindström G, Siebert J, Sivapalan M, Viney NR. 2009. Assessing the impact of land use change on hydrology by ensemble modeling (LUCHEM). I: Model intercomparison with current land use. Advances in Water Resources, 32(2): 129-146.

Culbertson AM, Martin J, Aloysius NR, Ludsin SA. Modeling the impacts of climate change on 21st century Maumee River water and nutrient discharge. In review.

Confesor R, Johnson L, Baker D. Shooting a moving target: A proposed critical source area paradigm for NW Ohio, in preparation.

Douglas-Mankin KR, Srinivasin R, Arnold JG. 2010. Soil and water assessment tool (SWAT) model: Current developments and applications. Transactions of the American Society of Agricultural and Biological Engineers, 53(5): 1423-1431.

Engel B, Storm D, White M, Arnold J, Arabi M. A hydrologic/water quality model application protocol. 2007. Journal of the American Water Resources Association, 43(5): 1223-1236.

Fales, M, Dell R, Herbert M, Doran PJ, Sowa SP. Making the leap from science to implementation: strategic agricultural conservation in the Saginaw Bay Watershed. Journal of Great Lakes Research. In press.

Gebremariam S, Martin J, DeMarchi C, Bosch N, Confesor R, Ludsin S. 2014. Comprehensive evaluation of three watershed models to predict Maumee River flow regime and discharge into Lake Erie. Environmental Modelling & Software, 61: 121-134.

Gildow M, Aloysius R, Gebremariam SY, Martin J. Fertilizer placement and application timing as strategies to reduce phosphorus loading to Lake Erie. In review.

GLWQA 2016. The United States and Canada adopt phosphorus load reduction targets to combat Lake Erie algal blooms. Available at: https://yosemite.epa.gov/opa/admpress.nsf/0/D6FB4CB50080797585257F610067D8BD.

International Joint Commission (IJC). 1988. Report on Modeling the Loading-Concentration Relationship for Critical Pollutants in the Great Lakes. IJC Great Lakes Water Quality Board, Toxic Substances Committee, Task Force on Toxic Chemical Loadings.

Kalcic MM, Kirchhoff C, Bosch N, Muenich RL, Murray M, Scavia D. Engaging Stakeholders to Define Feasible and Desirable Agricultural Conservation in Western Lake Erie Watersheds. In review.

Keitzer S, Ludsin SA, Sowa S, Annis G, Daggupati P, Froelich AM, Herbert ME, Johnson MV, Yen H, White MJ, Arnold JG, Sasson AM, Rewa CA. Thinking outside the lake: How might Lake Erie nutrient management benefit stream conservation in the watershed? In review.

Moriasi DN, Arnold JG, Van Liew MW, Binger RL, Harmel RD, Veith T. 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Transactions of the ASABE, 50(3): 885-900.

Muenich RL, Kalcic MM, Scavia D. Testing long-term agricultural nutrient management scenarios for the Maumee River Watershed. In review.

Murphy JC, Hirsch RM, Sprague LA. 2013. Nitrate in the Mississippi River and its tributaries, 1980–2010—An update: U.S. Geological Survey Scientific Investigations Report 2013–5169, 31 p.

Natural Resources Conservation Service (NRCS). 2016. U.S. effects of conservation practice adoption on cultivated cropland acres in Western Lake Erie Basin, 2003-06 and 2012. 120pp.

Obenour DR, Groneworld AD, Stow CA, Scavia D. 2014. Using a Bayesian hierarchical model to improve Lake Erie cyanobacteria bloom forecasts. Water Resources Research, 50: 7847-7860.

Robertson DM, Saad DA. 2011. Nutrient inputs to the Laurentian Great Lakes by source watershed estimated using SPARROW watershed models. Journal of the American Water Resources Association, 47(5): 1011-1033.

Robertson DM, Schwarz GE, Saad DA, Alexander RB. 2009. Incorporating uncertainty into the ranking of SPARROW model nutrient yields from Mississippi/Atchafalaya River Basin watersheds. Journal of the American Water Resources Association, 45(2): 534-549.

Scavia D, DePinto J. 2015. Annex 4 Ensemble Modeling Report. Available at: http://graham.umich.edu/scavia/wp-content/uploads/2014/11/Annex-4-Text-11-22-14.docx.

Scavia D, DePinto J, Bertani I A multi-model approach to evaluating target phosphorus loads for Lake Erie. In review.

Scavia D, Justić D, Bierman VJ. 2004. Reducing hypoxia in the Gulf of Mexico: Advice from three models. Estuaries, 27: 419-425.

Schwarz, GE, Hoos, AB, Alexander, RB, Smith, RA, 2006. The SPARROW surface water-quality model—Theory, application, and user documentation: U.S. Geol. Surv. Techniques and Methods, book 6, chap. B3.

Seiller G, Anctil F, Perrin C. 2012. Multimodel evaluation of twenty lumped hydrological models under contrasted climate conditions. Hydrology and Earth System Sciences, 16(4): 1171-1189.

Sowa SP, Herbert M, Cole L, et al. How much conservation is enough? Defining implementation goals for healthy fish communities. Journal of Great Lakes Research. In press.

Sprague LA, Hirsch RM, Aulenbach BT. 2011. Nitrate in the Mississippi River and its tributaries, 1980 to 2008: Are we making progress? Environmental Science & Technology, 45(17): 7209-7216.

Stow CA, Roessler C, Borsuk ME, Bowen JD, Reckhow KH. 2003. Comparison of estuarine water quality models for total maximum daily load development in Neuse River Estuary. Journal of Water Resources Planning and Management, 129 (4): 307-314.

Tebaldi C, Knutti R. 2007. The use of multi-model ensemble in probabilistic climate projections. Philosophical Transactions of the Royal Society A, 365(1857): 2053-2075.

United States Geological Survey (USGS). 2016. Summary of nitrogen, phosphorus, and suspended sediment loads and trends measured at the Chesapeake Bay nontidal network stations: Water year 2014 Update. Available at: http://cbrim.er.usgs.gov/summary.html.

Van Liew MW, Veith TL, Bosch DD, Arnold JG. 2007. Suitability of SWAT for the conservation effects assessment project: Comparison on USDA agricultural research service watersheds. Journal of Hydrological Engineering, 12(2): 173-189.

Vollmer-Sanders C, Allman A, Busdeker D, et al. Building partnerships to scale conservation: 4R nutrient stewardship certification program in the Lake Erie Watershed. Journal of Great Lakes Research. In press.

Velazquez JA, Schmid J, Ricard S, Muerth MJ, St-Denis BG, Minville M, Chaumont D, Caya D, Ludwig R, Turcotte R. 2013. An ensemble approach to assess hydrological models' contribution to uncertainties in the analysis of climate change impact on water resources. Hydrology and Earth System Sciences, 17: 565-578.

Wellen, C., Kamran-Disfani, A.R. and Arhonditsis, G.B., 2015. Evaluation of the current state of distributed watershed nutrient water quality modeling. Environmental Science & Technology, 49(6), pp.3278-3290.

Weller, D. E., B. Benham, M. Friedrichs, R. Najjar, M. Paolisso, P. Pascual, G. Shenk, and K. Sellner. 2013. Multiple Models for Management in the Chesapeake Bay STAC Publication Number 14-004, Edgewater, MD. 37 p.

Wilson R, Burnett L, Ritter T, Roe B, Howard G. 2013. Farmers, phosphorus and water quality: A descriptive report of beliefs, attitudes and practices in the Maumee Watershed of Northwest Ohio. Ohio State University, School of Environment & Natural Resources. Available at: http://ohioseagrant.osu.edu/archive/maumeebay/docs/farmers-phosphorus-and-water-quality.pdf.

Yen H, Wang X, Fontane DG, Arabi M, Harmel RD. 2014a. A framework for propagation of uncertainty contributed by input data, parameterization, model structure, and calibration/validation data in watershed modeling. Environmental Modelling and Software, 54, pp. 211-221.

Yen H, Bailey RT, Arabi M, Ahmadi M, White MJ, Arnold JG. 2014b. The role of interior watershed processes in improving parameter estimation and performance of watershed models. Journal of Environmental Quality, 43(5), pp. 1601-1613.

Yue X, Mickley LJ, Logan JA, Kaplan JO. 2013. Ensemble projections of wildfire activity and carbonaceous aerosol concentrations over the western United States in the mid-21st century. Atmospheric Environment, 77: 767–780.

Appendices

Appendix A1: Explanation of the estimates of sources and delivery of P in the Maumee River Basin

Appendix A2: Details on the differences among SWAT models

Appendix A3: SPARROW Model Details

Appendix A4: Details on baseline homogenized meteorological data

Appendix A5: Details on baseline point source data

Appendix A6: Details on observed data used for validating the models

Appendix A7: Single-practice scenario results

Appendix A8: Bundled scenarios details

Appendix A9: Potential Hotspot Identification Methods

Appendix A10: Baseline validation results

A1. Explanation of the estimates of sources and delivery of P in the Maumee River Basin

A ballpark estimate of the sources and delivery of P in the Maumee watershed provided context for the project, and was entirely separate from the watershed modeling and main results of this report. The estimate considered all major sources of P in the watershed, but did not include atmospheric deposition of P, which may account for 2% of TP and 3% DRP in the Maumee River (Maccoux et al. in review); atmospheric sources likely originate from wind erosion from the watershed, which is predominantly managed in agricultural lands. The purpose of this work was to determine the maximum potential contribution from point sources and non-farm fertilizers, and so it is a conservative estimate of the agricultural contribution of P to Lake Erie.

Total input and potential delivery of all known phosphorus sources to the Maumee watershed was estimated, including inputs from point sources, septic systems, nonfarm fertilizers, farm fertilizers, and manure (see Report text, Figure 1).

Point Sources

Point source input to the watershed was estimated as the point source data already compiled for Baseline models in Appendix A5, downloaded from EPA's Discharge Monitoring Report (DMR) Pollutant Loading Tool (https://cfpub.epa.gov/dmr/). Monthly average phosphorus loads were summed to calculate a total annual load of 141,590 kg/year. This estimate does not include the Toledo wastewater treatment plant, as it discharges to the mouth of the river in Maumee Bay and therefore was not included in the Baseline models. A separate estimate of the Toledo plant was conducted to ensure the approach erred on the side of over-estimating the contributions from point sources in the watershed. The total annual phosphorus load for the Toledo treatment plant was estimated to be 54,430 kg/year, again derived from the EPA's DMR Pollutant Loading Tool.

Point source delivery to Lake Erie was estimated as the total point source inputs in the watershed, thus assuming no in-stream processing of nutrients. While this is in practice a poor assumption, the purpose of this pie chart is to show a *conservative* estimate for agricultural sources, and therefore we represent the worst case for point sources.

Septic Systems

Septic system input to the watershed was estimated as follows: Data was retrieved from the National Environmental Services Center (NESC) (http://www.nesc.wvu.edu/) for counties in Michigan (http://www.nesc.wvu.edu/septic_idb/michigan.htm) and Ohio (http://www.nesc.wvu.edu/septic_idb/ohio.htm) in the Maumee watershed, including total population, total housing units, and the number of people on septic tank systems per county for 1990. The same level of data was not available for Indiana, and so the number of people on septic was estimated from a relationship between county population and percentage of population on septic found in Ohio and Michigan counties (Figure A1.1). A similar calculation was performed by using population estimates for all of the Maumee counties for 2014. The sum of all people on septic systems for all of the Maumee Basin counties (1990 & 2014) was then calculated and multiplied by the total amount of phosphorus produced by one person per year to determine the total amount of phosphorus produced in the Maumee River basin counties. County populations were derived from the U.S. Census Bureau 2014 National Population Projections (https://www.census.gov/population/projections/data/national/2014.html). The estimated total amount of phosphorus (TP) excreted in urine and feces by one person was estimated on average to be 0.582 kg/person-year (Mihelcic et al. 2011). For 1990, the estimated TP contributed to the system by septic systems was 255,443 kg/year and for 2014 this number was slightly higher at 256,350 kg/year. Many of the counties are not fully inside the Maumee watershed, so the result from each county was weighted by its fraction within the watershed for a final input for 2014 of 167,010 kg/year.

Septic delivery to Lake Erie was assumed to be 39% of the total septic input, or 65,130 kg/year phosphorus, as the septic failure rate for that region is approximately 39% according to a 2013 Ohio Department of Health Report on septic systems and failures (https://www.odh.ohio.gov/~/media/ODH/ASSETS/Files/eh/STS/2012HSTSSystemsandFailures.pdf). Similar to point sources, this was a worst case scenario for septic systems; it is highly improbable that all the phosphorus from failing septic systems would reach the outlet of the Maumee without subsequent retention or treatment by the soil or stream network.

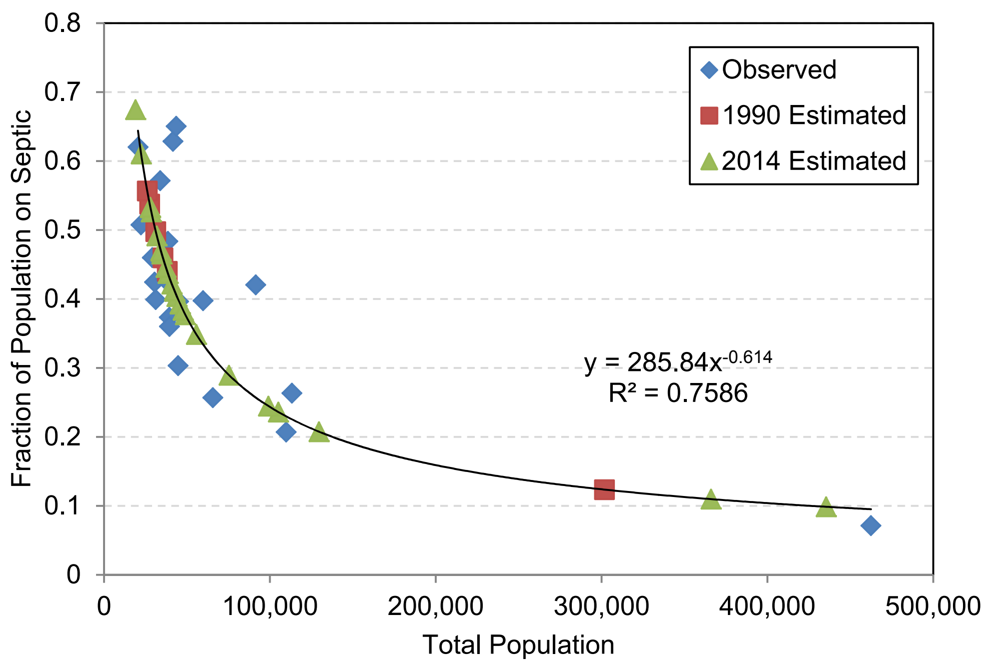


Figure A1.1: Relationship between fraction of population on septic systems and the total population of a county for MI and OH. The 1990 estimate (red squares) are for counties in Indiana. The 2014 estimates (green triangles) are for all counties in the MRW. The blue diamonds are observed values from NESC.

Fertilizers & Manures

Inputs from farm fertilizers, nonfarm fertilizers, and manure to the watershed were estimated from a USGS report including county-level estimates of annual farm and nonfarm fertilizer sales for 1987-2001 as well as manure production for 1992-1997 (Ruddy et al. 2006). Data from multiple years were averaged as there was no obvious temporal trend. For manure, both confined and unconfined manure were combined. County values were weighted by their fraction within the Maumee watershed and summed. The average annual phosphorus input from nonfarm fertilizer was 129,710 kg/year, the input from farm fertilizer was 19,280,780 kg/year, and the input from manure production was 6,024,080 kg/year.

Nonfarm fertilizer delivery to Lake Erie was estimated as all nonfarm fertilizer input, assuming no fertilizers remain on the land or are processed in the river. As with the point sources and septic systems, this is an intentional over-estimation of the non-farm fertilizer contribution so that the contribution from farm fertilizers and manures would be conservative.

Farm fertilizer and manure delivery to Lake Erie was calculated as the remaining load not accounted for by all other fractions. The average annual total phosphorus load to Lake Erie from 2002-2014 was estimated from Maccoux et al. (in review), and the contribution from point sources, septic systems, and nonfarm fertilizers was subtracted from this load. The remainder is a conservative estimate of farm fertilizer contribution, one unaccounted for source may be legacy phosphorus from farming activities on the land and stored in the stream system. These are average conditions; on a high-flow year the farm contribution would be an even greater total load and percentage of the load to Lake Erie.

References

Mihelcic JR, Fry LLM, Shaw R. 2011. Global potential of phosphorus recovery from human urine and feces. Chemosphere, 84(6): 832-839.

> *Phosphorus excreted per person annually was determined from Table 1 by dividing the total phosphorus in human excreta per year (417,708 metric tons) by the total population for "Developed" countries (718,279,000).*

Ruddy BC, Lorenz DL, Mueller, DK. 2006. County-level estimates of nutrient inputs to the land surface of the conterminous United States, 1982-2001. USGS Scientific Investigations Report 2006-5012. Available online at: http://pubs.usgs.gov/sir/2006/5012/.

Maccoux MJ, Dove A, Backus SM, Dolan DM. (in review). Total and soluble reactive phosphorus loadings to Lake Erie. Journal of Great Lakes Research.

A2. Details on the differences among SWAT models

Although the five SWAT models all use the same base model, there are a multitude of differences between them that make each in essence different models. The two tables (A2.1 and A2.2) describe a few of these differences and may allow for further interpretation of differences in results across the models.

Table A2.1: Comparison of individual modeling decisions and inputs.

Aspect of SWAT Modeling	Modeling Decision	Decision Options	Models				
			HU	LT	OSU	TAMU	UM
Model/Sub-Model Algorithms	Model Version	Rev. 635-modified†		x	x	x	x
		Rev. 637-modified†	x				
	Tile Drain Routine	Old (SWAT_TDRAIN)				x	
		New (SWAT_HKdc)	x	x	x		x
	Water Table Routine	Old	x	x	x	x	
		New					x
	In-Stream Processes	On (QUAL2E)	x		x	x	x
		On, modified‡		x			
	Soil P Model	Old	x			x	
		New		x	x		x
	Evapotranspiration Method	Penman-Monteith	x		x	x	x
		Hargreaves		x			
Model Inputs	Land Use Data	NLCD 2001		x			
		NLCD 2006					x
		CDL 2007			x		
		CDL 2010-2011				x	
		CDL 2009-2012	x				
	Elevation Model	NED 10m	x				
		NED 30m		x	x	x	x
	Soils Data	SSURGO	x		x	x	x
		STATSGO		x			
	Climate Inputs*	NOAA NCDC - precipitation and temperature	x	x	x	x	x
		Simulated solar radiation, wind, relative humidity	x	x	x	x	x
	Point Source Inputs*	Measured data from EPA DMR; aggregated to average monthly	x	x	x	x	x
Spatial Discretization	HRU Thresholds	LU-Soil-Slope: 0/10/0					x
		LU-Soil-Slope: 200 ha/800 ha/800 ha			x		
		LU-Soil-Slope: 5/10/0		x		x	
		LU-Soil-Slope: 50/25/0	x				
	# Subbasins	*Calculation after model setup*	265	203	252	391	358
	Average HRU Area (ha)	*Calculation after model setup*	107	727	800	72	169

*Data homogenized for this project.
† SWAT versions were modified to fix a bug where soluble P was not properly moving through subsurface drains.
‡watqual3 routine is an adaption LimnoTech developed based on the paper: White MJ, Storm DE, Mittelstet A, Busteed PR, Haggard BE, Rossi C. 2014. Development and Testing of an In-Stream Phosphorus Cycling Model for the Soil and Water Assessment Tool. Journal of Environmental Quality, 215.

Table A2.1, continued: Comparison of individual modeling decisions and inputs.

Aspect of SWAT Modeling	Modeling Decision	Decision Options	Models: HU	LT	OSU	TAMU	UM
Model Parameterization & Measured Data	Methods for Assessing Model Performance	R^2	x		x		x
		NSE	x	x	x	x	x
		PBIAS		x	x	x	x
	Variables Model Performance Was Assessed For	Streamflow	x	x	x	x	x
		Total Phosphorus	x	x	x	x	x
		Dissolved Reactive Phosphorus	x	x	x	x	x
		Total Nitrogen	x	x		x	x
		Nitrate	x	x		x	x
		Sediment	x		x	x	x
	Additional Calibration Checks	Crop Yields	x	x	x	x	x
		Tile Flow	x	x	x	x	x
		Field Losses		x			
		Nutrient Loss via Tile Drains		x	x	x	x
	Calibration Time Period	2001-2005					x
		2000-2009			x		
		1998-2010		x			
		2009-2012	x				
		1990-1999				x	
	Spatial Extent of Calibration	At Waterville only	x		x	x	x
		At Waterville, Blanchard and Tiffin		x			
	Method to Fill in Missing Data	LOADEST for everything except DRP; Obenour et al. (2014) method for DRP				x	
		Model is calibrated only to observed data; missing data not included in calibration	x	x	x		x

Reference:
Obenour DR, Groneworld AD, Stow CA, and Scavia D. 2014. Using a Bayesian hierarchical model to improve Lake Erie cyanobacteria bloom forecasts. Water Resources Research, 50: 7847-7860.

Table A2.1, continued: Comparison of individual modeling decisions and inputs.

Aspect of SWAT Modeling	Modeling Decision	Decision Options	Models				
			HU	LT	OSU	TAMU	UM
Land Management Operations	Fertilizer Applications	Estimated from county fertilizer sales data from 2002					x
		Estimated based on maintenance application from Tri-State Standards	x		x		
		Aggregated inputs from USDA-ARS NHDPlus SWAT model (Daggupati et al. 2015)		x			
		Estimated from Ag Census yield and Fertilizer Use data 1990-2010				x	
	Manure Applications	Estimated from Ag Census					x
		Aggregated inputs from USDA-ARS NHDPlus SWAT model (Daggupati et al. 2015)		x			
		Not included	x		x	x	
	Crop Rotations (C = Corn, S = Soybean, W = Winter Wheat, H = Hay)	CS	x	x	x	x	x
		CSS	x		x		x
		CSW	x		x	x	
		CWS		x			
		CSWCSSW					x
		CSWH			x		
		SS	x	x	x	x	
		CC	x	x	x	x	
	Tillage	Estimated from CTIC					x
		Estimated from USDA/OSU Extension consultation			x		
		Estimated according to crop planted	x				
		Estimated based on modified RUSLE2		x		x	
	Tile Drainage	All agricultural lands with somewhat poorly, poorly, or very poorly drained soils					x
		C,S,W HRU's with poorly or very poorly drained soils			x		
		AGRR or HAY lands with hydrologic group C or D soils		x			
		Ag lands with less than or equal to 3% slope	x				
		Ag lands with <1% slope				x	

Reference:

Daggupati P, Yen H, White MJ, Srinivasan R, Arnold JG, Keitzer SC, Sowa SP. 2015. Impact of model development decisions on hydrological processes and streamflow simulations in West Lake Erie basin. Hydrological Processes, 29(26), pp. 5307-5320.

Table A2.2: Individual model parameter values. Values highlighted in gray for a given model indicate that the parameter was actually changed from its default value for that model. Values not highlighted were left at default (DF); since different model versions may have different model defaults, the basin-level default values were included. *Indicates value was only changed on tile-drained lands. †Indicates value was changed by a percentage, and is therefore not an absolute value for the parameter. Some values have been rounded off for presentation purposes. NA indicates the parameter was not applicable to the model given the set of sub-routines activated.

Parameters that turn sub-routines on or off									
					Final or Calibrated Value				
Parameter	**File**	**Spatial Level**	**Description**	**Range**	**HU**	**LT**	**OSU**	**TAMU**	**UM**
ICN	.bsn	Watershed	Daily curve number calculation method: 0-calculate daily CN value as a function of soil moisture; 1-calculate daily CN value as a function of plant evapotranspiration	0/1	0	0	0	1	0
IRTE	.bsn	Watershed	Channel water routing method; 0=variable travel-time; 1=Muskingum	0/1	0	0	1	0	0
ISMAX	.bsn	Watershed	Maximum depressional storage flag, 0 = static stmaxd from .sdr	0/1	0	0	0	NA	1
ITDRN	.bsn	Watershed	Tile drainage equations flag; 1=SWAT_HKdc routine using DRAINMOD; 0=SWAT_TDRAIN method.	0/1	1	1	1	0	1
IWQ	.bsn	Watershed	In-stream water quality model: 0-do not simulate nutrient transformations in stream; 1-activate simulation of in-stream nutrient transformations using QUAL2E; 2-watqual2 simulation; 3-watqual3‡.	0/1	1	3	1	1	1
IWTDN	.bsn	Watershed	Water table depth algorithms flag	0/1	1	0	0	0	1
SOL_P_MODEL$^{\Delta}$	.bsn	Watershed	Soil phosphorus sub-routine: 0=new model; 1=old model	0/1	1	0	0	1	0

‡watqual3 routine is an adaption LimnoTech developed based on: White MJ, Storm DE, Mittelstet A, Busteed PR, Haggard BE, Rossi C. 2014. Development and Testing of an In-Stream Phosphorus Cycling Model for the Soil and Water Assessment Tool. Journal of Environmental Quality, 215.

Table A2.2, continued: Individual model parameter values. Values highlighted in gray for a given model indicate that the parameter was actually changed from its default value for that model. Values not highlighted were left at default (DF); since different model versions may have different model defaults, the basin-level default values were included. *Indicates value was only changed on tile-drained lands. †Indicates value was changed by a percentage, and is therefore not an absolute value for the parameter. Some values have been rounded off for presentation purposes. NA indicates the parameter was not applicable to the model given the set of sub-routines activated.

Parameters that were calibrated in at least one model									
					Final or Calibrated Value				
Parameter	**File**	**Spatial Level**	**Description**	**Range**	**HU**	**LT**	**OSU**	**TAMU**	**UM**
ADJ_PKR	.bsn	Watershed	Peak rate adjustment factor	0.5-1.5	1.474	0	1	1	1
ALPHA_BF	.gw	HRU	Baseflow recession constant	0.1-0.99	0.937	0.254	DF	DF	DF
ANION_ EXCL	.sol	HRU	Fraction of soil pore space from which anions are excluded	0-1	DF	DF	0.5	DF	0.1
BC1	.swq	Subbasin	Biological oxidation rate of NH4 to NO2 in the reach at 20° (1/day)	0.1-1	DF	0.36	0.55	0.55	0.1
BC3	.swq	Subbasin	Hydrolysis rate of organic N to NH4 in the reach at 20° (1/day)	0.2-0.4	DF	DF	DF	DF	0.02
BC4	.swq	Subbasin	Mineralization rate of organic P to DRP in the reach at 20° (1/day)	0.01-0.7	0.012	0.02	0.05	0.004	0.01
BIOMIX	.mgt	HRU	Biological mixing efficiency	NA	DF	0.2	0.75	DF	0.3
CANMX	.hru	HRU	Maximum canopy storage (mm H2O)	NA	DF	5.732	DF	DF	DF
CDN	.bsn	Watershed	Rate coefficient for dentirification	0-3	1.4	0.5	1.4	.181	1.4
CH_COV1	.rte	Subbasin	Channel cover factor 1	0-1	DF	0.048	0	0.037	0.5
CH_COV2	.rte	Subbasin	Channel cover factor 2	0-1	DF	0.048	0	0.219	0.5
CH_K1	.sub	Subbasin	Effective hydraulic conductivity (mm/hr)	0.025-25	9.811	DF	DF	DF	DF
CH_K2	.rte	Subbasin	Effective hydraulic conductivity of channel (mm/hr)	0.025-25	13.65	DF	DF	DF	DF
CH_N1	.sub	Subbasin	Manning's roughness for tributary channels	0-0.15	0.117	DF	0.014	0.014	0.025
CH_N2	.rte	Subbasin	Manning's roughness for the main channel	0-0.15	0.016-0.149	0.057	0.014	0.005	0.035

Table A2.2, continued: Individual model parameter values. Values highlighted in gray for a given model indicate that the parameter was actually changed from its default value for that model. Values not highlighted were left at default (DF); since different model versions may have different model defaults, the basin-level default values were included. *Indicates value was only changed on tile-drained lands. †Indicates value was changed by a percentage, and is therefore not an absolute value for the parameter. Some values have been rounded off for presentation purposes. NA indicates the parameter was not applicable to the model given the set of sub-routines activated.

Parameters that were calibrated in at least one model									
					Final or Calibrated Value				
Parameter	**File**	**Spatial Level**	**Description**	**Range**	**HU**	**LT**	**OSU**	**TAMU**	**UM**
CN2	.mgt	HRU	Initial SCS moisture condition II curve number	0.75-1.25†	28.1-99.9	30-95	DF	DF	DF
CNOP	.mgt	HRU	SCS runoff curve number for moisture condition II	NA	DF	75-89	DF	DF	DF
DDRAIN	.mgt	HRU	Depth to subsurface tile drain (mm)	0-6000	915*	1000*	900*	~1220*	1000*
DEP_IMP	.hru	HRU	Depth to the impervious layer in the soil (mm)	0-6000	2500*	2500*	3370*	2381*	1500*
DRAIN_CO	.sdr	HRU	Daily drainage coefficient (mm/day)	10-51	DF	12.7	10	NA	25
EPCO	.bsn	Watershed	Plant uptake compensation factor.	0.01-1.0	1.0	0.638	1.0	1.0	1.0
ERORGN	.hru	HRU	Nitrogen enrichment ratio for loading with sediment, 0 allows model to calculate value	NA	DF	1.1	DF	DF	DF
ERORGP	.hru	HRU	Phosphorus enrichment ratio for loading with sediment, 0 allows model to calculate value	NA	DF	1-1.2	DF	DF	DF
ESCO	.bsn, .hru	Watershed HRU	Soil evaporation compensation factor	0.01-1	0.78^{bsn}	1^{bsn}	0.99^{hru}	0.967^{bsn}	1^{bsn}
GDRAIN	.mgt	HRU	Drain tile lag time (hours)	NA	NA	NA	NA	24	NA
GW_DELAY	.gw	HRU	Delay time for aquifer recharge (days)	NA	3.747	DF	DF	DF	DF
GWQMN	.gw	HRU	Threshold water level in shallow aquifer for base flow (mm H2O)	NA	32.41	447.6	DF	DF	DF

Table A2.2, continued: Individual model parameter values. Values highlighted in gray for a given model indicate that the parameter was actually changed from its default value for that model. Values not highlighted were left at default (DF); since different model versions may have different model defaults, the basin-level default values were included. *Indicates value was only changed on tile-drained lands. †Indicates value was changed by a percentage, and is therefore not an absolute value for the parameter. Some values have been rounded off for presentation purposes. NA indicates the parameter was not applicable to the model given the set of sub-routines activated.

Parameters that were calibrated in at least one model									
					Final or Calibrated Value				
Parameter	**File**	**Spatial Level**	**Description**	**Range**	**HU**	**LT**	**OSU**	**TAMU**	**UM**
GW_REVAP	.gw	HRU	Revap coefficient	0.02-2	1.41	DF	DF	DF	DF
HRU_SLP	.hru	HRU	Average slope steepness (m/m)	0.75-1.25†	0.97†	DF	DF	DF	DF
IFLOD1R	.res	Subbasin	Beginning month of non-flood season	1-12	DF	12	DF	DF	DF
IFLOD2R	.res	Subbasin	Ending month of non-flood season	1-12	DF	1	DF	DF	DF
LATKSATF	.sdr	HRU	Lateral soil hydraulic conductivity in tile-drained fields as multiple of original soil conductivity value	0.01-4	DF	2-4	1	NA	1
NDTARGR	.res	Subbasin	Number of days to reach target storage from current reservoir storage	NA	DF	5	DF	DF	DF
NPERCO	.bsn	Watershed	Nitrate percolation coefficient	0.01-1	0.391	0.5	0.2	0.394	0.4
OVN	.hru	HRU	Manning's "n" value for overland flow	0.008-0.5	0.437	DF	DF	DF	DF
PHOSKD	.bsn	Watershed	Phosphorus soil partitioning coefficient (m^3/Mg)	80-350	326.9	175	200	422.5	175
PPERCO	.bsn	Watershed	Phosphorus percolation coefficient (m^3/Mg)	10-17.5	10	10	10	17.16	10
PSP	.bsn	Watershed	Phosphorus availability index	0.2-0.6	0.231	0.4	0.4	0.215	0.4
R2ADJ	.hru	HRU	Curve number adjustment for increasing infiltration in non-draining soils	0-3	DF	1.75-3.0	1	DF	8*
RE	.sdr	HRU	Effective radius of drains (mm)	3-40	DF	10*	DF	NA	DF

Table A2.2, continued: Individual model parameter values. Values highlighted in gray for a given model indicate that the parameter was actually changed from its default value for that model. Values not highlighted were left at default (DF); since different model versions may have different model defaults, the basin-level default values were included. *Indicates value was only changed on tile-drained lands. †Indicates value was changed by a percentage, and is therefore not an absolute value for the parameter. Some values have been rounded off for presentation purposes. NA indicates the parameter was not applicable to the model given the set of sub-routines activated.

Parameters that were calibrated in at least one model									
					Final or Calibrated Value				
Parameter	**File**	**Spatial Level**	**Description**	**Range**	**HU**	**LT**	**OSU**	**TAMU**	**UM**
REVAPMN	.gw	HRU	Threshold water level level in shallow aquifer for revap (mm H_2O)	NA	97.06	388.6	DF	DF	DF
RS2	.swq	Subbasin	Benthic source rate for DRP in the reach at 20° (mg P/m^2-d)	NA	DF	0.05	0.05	0.022	0.01
RS3	.swq	Subbasin	Benthic source rate for ammonium in the reach at 20° (mgNH4-N/m2/d)	NA	DF	0.5	0.5	DF	1
RS4	.swq	Subbasin	Organic N settling rate in the reach at 20° (1/day)	0.001-0.1	DF	0.05	0.05	DF	0.001
RS5	.swq	Subbasin	Local settling rate for organic phosphorus mineralization at 20° (day^{-1})	0.001-0.1	DF	0.07	0.05	DF	0.05
SDNCO	.bsn	Watershed	Threshold value of nutrient cycling water factor for denitrification to occur	0.75-1.4	1.005	1	1.1	1.041	1.1
SDRAIN	.sdr	HRU	Tile drain spacing (mm)	7,600-30,000	DF*	13720*	15000*	NA	15000*
SFTMP	.bsn	Watershed	Mean air temperature at which precipitation is equally likely to be rain as snow/freezing rain (°C)	-5-5	-1.51	1	1	1	-2
SHALLST	.gw	HRU	Initial depth of water in the shallow aquifer (mm H2O)	NA	DF	500	DF	DF	DF
SLSUBSN	.hru	HRU	Average slope length	0.75-1.25	0.97†	DF	DF	DF	DF
SMFMN	.bsn	Watershed	Minimum snow melt factor (mm H_2O/day-°C)	1.4-6.9	3.547	3	4.5	4.5	2
SMFMX	.bsn	Watershed	Maximum snow melt factor (mm H_2O/day-°C)	1.4-6.9	6.027	4.5	4.5	2.5	2
SMTMP	.bsn	Watershed	Threshold temperature for snowmelt (°C)	-5-5	1.611	0.5	0.5	2.5	-2

Table A2.2, continued: Individual model parameter values. Values highlighted in gray for a given model indicate that the parameter was actually changed from its default value for that model. Values not highlighted were left at default (DF); since different model versions may have different model defaults, the basin-level default values were included. *Indicates value was only changed on tile-drained lands. †Indicates value was changed by a percentage, and is therefore not an absolute value for the parameter. Some values have been rounded off for presentation purposes. NA indicates the parameter was not applicable to the model given the set of sub-routines activated.

Parameters that were calibrated in at least one model									
					Final or Calibrated Value				
Parameter	**File**	**Spatial Level**	**Description**	**Range**	**HU**	**LT**	**OSU**	**TAMU**	**UM**
SOL_AWC	.sol	HRU	Available water capacity	0.75-1.25	0.96†	DF	DF	DF	DF
SOL_CRK	.sol	HRU	Potential crack volume for soil profile	0-1	DF	DF	DF	0.11	0.45
SOL_K	.sol	HRU	Saturated hydraulic conductivity (mm/hr)	0.75-1.25	0.92†	DF	DF	DF	DF
SOL_ORGP	.chm	HRU	Initial humic organic phosphorus in soil layer (mg/kg or ppm)	50-250	94.906	DF	DF	DF	DF
SOL_SOLP	.chm	HRU	Initial labile P in the soil layer (mg labile P/kg soil)	5-100	7.002	DF	10	34	1
SPCON	.bsn	Watershed	Parameter drives the maximum concentration of sediment the river can route	0.0001-0.01	1e-4	1e-3	1e-4	2.3e-3	2.7e-4
SURLAG	.bsn	Watershed	Surface runoff lag coefficient	NA	1.08	2.872	4	0.023	1
TDRAIN	.mgt	HRU	Time to drain soil to field capacity (hours)	NA	NA	NA	NA	48	NA
TIMP	.bsn	Watershed	Snow pack temperature lag	0.01-1	0.13	0.06	1	1	0.05
USLE_C	crop.dat	By land-use	Minimum value for the cover and management factor for the land cover	0.75-1.25	1.21†	DF	DF	DF	DF
USLE_K	.sol	HRU	USLE soil erodibility factor (0.013 metric ton m^2-hr/m^3- metric ton cm)	0.75-1.25	0.887†	DF	DF	DF	DF
USLE_P	.mgt	HRU	USLE support practice factor	0.50-1.25	1.078†	0.6-1.0	DF	DF	DF
VCRIT	.bsn	Watershed	Critical velocity at which a river will resuspend sediments	NA	5	0	5	5	1

$^{\Delta}$SWAT 2012 revision 635 indicate in basins.bsn that 1 is the new soil phosphorus model; however, examination of the source code followed by confirmation from Nancy Sammons (in a post to the SWAT-user group on 2/26/2014) confirms that setting this parameter equal to 0 will run the new soil phosphorus sub-routine.

A3. SPARROW Model Details

The SPARROW model was also used in this study; details on its set-up are included in Table A3.1 and methods applied specifically to update the Maumee River Watershed for this study are described in detail below. This adjustment replaces the original SPARROW delivered load estimates from point sources with 2011 point source inputs and then rescales the remaining portion of the original SPARROW load estimate to match the 2011 monitored load at Waterville, OH.

<u>Methods to adjust the SPARROW Model for the Maumee River Watershed</u>

1. Determine the original, local point source delivery ratio from each SPARROW catchment to the local stream (based on the measured point source input to a catchment and the simulated incremental point source contribution from the catchment) and downstream point-source delivery ratio from the SPARROW catchment to the basin outlet (based on the measured point source input to a catchment and the simulated delivered incremental point source contribution from that catchment to the basin outlet). Delivered contributions represent the amount of the load leaving a reach that is not attenuated or removed by natural processes during downstream transport to a specified downstream target reach. In this example, the downstream target reach is the Maumee River at the gaging station at Waterville, Ohio. The original point source delivery rates to the stream and delivered to the basin outlet will be applied to the 2011 point source inputs.
2. Determine the original non-point source contribution from the entire basin by subtracting total delivered point source contribution from SPARROW from the total delivered load from the original SPARROW model.
3. Replace original point source inputs with 2011 point source inputs into each catchment; these were identical to the SWAT modeling inputs.
4. Estimate new point source contributions to the incremental and delivered incremental loads from each SPARROW catchment. This is done by multiplying the new point source inputs to each catchment by the local point source delivery ratio and the downstream point-source delivery ratio identified in Step 1. Note these ratio values can be more than 1.0.
5. Obtain the 2011 measured load at the basin outlet (these data are the same as that used by the SWAT modelers for validation and are the data provided by Heidelberg University at the station near Waterville, Ohio). In this application, the measured load at the basin outlet is the mean annual phosphorus load detrended to 2011.
6. Estimate the new non-point source load from the entire basin by subtracting the sum of the new delivered incremental point source loads from Step 4 from the measured basin load from Step 5.
7. Compute a basin-wide new non-point source load adjustment factor (New non-point load (Step 6) / Original SPARROW non-point load (Step 2)).
8. Adjust all of the original non-point source contributions from the original SPARROW model by the new non-point adjustment factor for each catchment.
9. Sum the incremental and delivered incremental loads, for each catchment, for each source, for the entire basin, to determine the updated source contributions based on 2011 point source inputs.

Methods to adjust SPARROW Model results to reflect new fertilizer inputs

1. Adjust the incremental and delivered incremental point and non-point (by source) loads, for each catchment, (from Steps 4 and 8 above, respectively), by the percentage provided. In this case it is reducing the fertilizer inputs by 50%.
2. Sum the incremental and delivered incremental loads, for each catchment, for each source, for the entire basin, to determine the updated catchment loads and yields, and the delivered catchment loads and yields.

Table A3.1: Data requirements and specifications for data used in the MRB3 SPARROW models. Obtained from Robertson and Saad (2011).

Category	Requirements/Specifications
Water-quality sites	
Time period for data	10/1/1970 - 9/30/2007
Time period covered by water-quality data	> 2 years
Data near 2002 base year	Data within 2 years of 2002 if < 5 years of data; data within 7 years of 2002 if > 5 years of data
Total number of samples	>25 samples
Total number of uncensored sample values	>25 samples
Total samples in each of four seasons	>1 sample for each season (winter: Dec.-Feb.; spring: Mar.-May; summer: June-Aug.; fall: Sept-Nov.)
Location of site	On enhanced RF1 stream coverage
Coinciding stream gage	
Time period for data	10/1/1970 - 9/30/2006
Water quality and flow data overlap	> 2 years
Drainage area ratio between water quality site and gaged site	0.5 - 2.0
Proximity between water quality site and gaged site	< 40 km
Proximity between water quality site and gaged site for large streams (>260 km^2)	Must be on the same stream network
Load Computations	
Program for load computation	Fluxmaster (Schwarz et al. 2006)
Variables included in Fluxmaster	logarithm of flow, sine, cosine, decimal time
Time period of data used in Fluxmaster calibration	10/1/1970 - 9/30/2007
Time period for load computation	10/1/1970 - 9/30/2006
Annual load computation period	Water year 10/1 - 9/30
Detrended to which year (base year)	2002
Point Sources	
Point sources not included for the following Standard Industrial Classification (SIC) codes	1389- Oil and gas injection wells; 3312, 3479, 3339 - Steel; and 4961- steam.
Model Calibration and Accumulation Procedures	
SPARROW version	V2_9
Coefficient Estimation	Nonlinear least square regression (NLLSR)
Confidence limits on coefficients	Compute coeffcients with NLLSR followed by application of Make_coef_ci.sas code.
Robustness of coefficients	200 nonparametric bootstrap iterations
Accumulation at a HUC8 scale	Not corrected for biases. Accumulated with Custom_predict_accumulator.sas; Confidence intervals computed with Sparrow_custom_predict.sas using 200 iterations of parametric bootstraps.
Accumulation for all tributaries > 150 km^2	Corrected for biases. Accumulated with Custom_predict_accumulator.sas; Confidence intervals computed with Sparrow_custom_predict.sas using 200 iterations of parametric bootstraps.
Accumulation for each Great Lake	Corrected for biases.Accumulated with Custom_predict_accumulator.sas; Confidence intervals computed with Sparrow_custom_predict.sas using 200 iterations of parametric bootstraps.

References

Robertson DM, Saad DA. 2011. Nutrient inputs to the Laurentian Great Lakes by source watershed estimated using SPARROW watershed models. Journal of the American Water Resources Association, 47(5): 1011-1033.

A4. Details on baseline homogenized meteorological data

This section describes the source of common meteorological data used in the Baseline models. NOAA Global Historical Climatology Network -Daily (GHCN-DAILY) data were retrieved (Menne et al. 2012) and Michigan, Ohio, and Indiana stations were extracted from the larger dataset. Stations within a bounding latitude and longitude box of the Western Lake Erie Basin were further extracted (Figure A4.1).

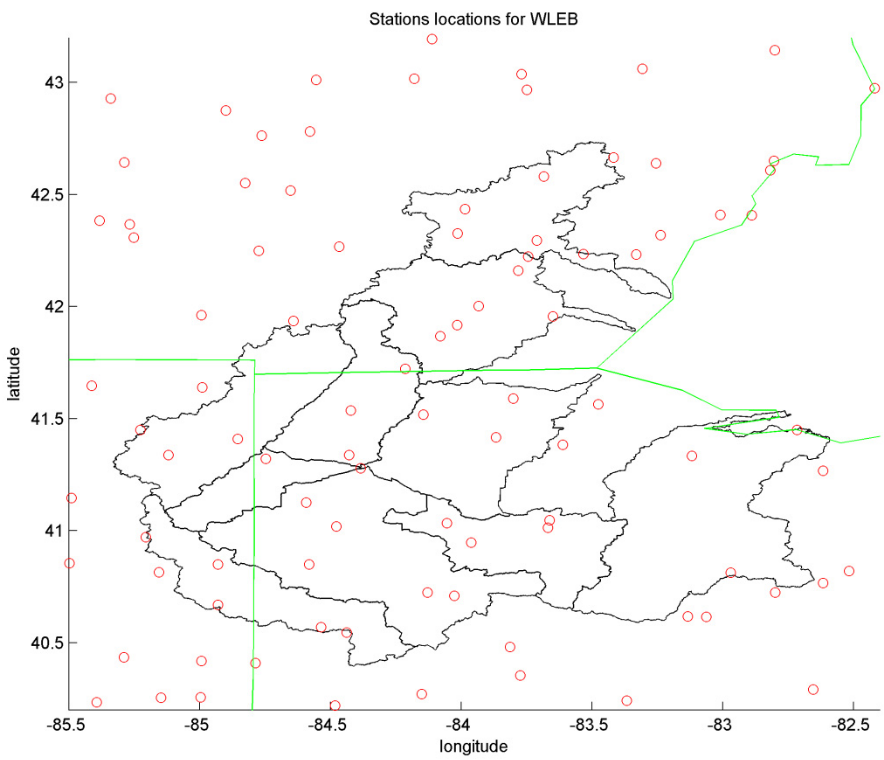


Figure A4.1: Location of NOAA GHCN stations in the Western Lake Erie Basin.

Data outliers were flagged as likely erroneous and removed if the daily temperature at a given station was more than +/- 20° from the daily station average or if the daily precipitation exceeded 300mm (Table A3.1). In the event that a station was an outlier or the original dataset was missing precipitation or temperature information, the value from the nearest station was used to fill in; if that station also had missing data the next closest station would be considered until a value was found. The resulting average meteorological data driving the baseline models from 2005 - 2014 (including a model spin-up from 2000 - 2004) are shown in Figures A4.2-A4.7.

Table A4.1: Outliers removed from WLEB stations in the time period 2000 - 2014. There were no precipitation outliers during this time period.

- TMIN outlier of -17.2 deg C found in station number US1INAD0001 on date 5/8/2007.
- TMIN outlier of -21.1 deg C found in station number US1INAL0017 on date 12/30/2004.
- TMIN outlier of -37.2 deg C found in station number US1INAL0019 on date 12/13/2000.
- TMIN outlier of -33.3 deg C found in station number US1INAL0019 on date 12/14/2000.
- TMAX outlier of 0.6 deg C found in station number US1INAL0020 on date 4/15/2000.
- TMIN outlier of -24.4 deg C found in station number US1INAL0020 on date 2/14/2005.
- TMIN outlier of -13.3 deg C found in station number US1INAL0020 on date 5/23/2005.
- TMIN outlier of -3.9 deg C found in station number US1INAL0042 on date 10/7/2007.
- TMAX outlier of -13.3 deg C found in station number US1INAL0051 on date 10/24/2003.
- TMIN outlier of -17.8 deg C found in station number US1INAL0053 on date 4/30/2002.
- TMAX outlier of 34.4 deg C found in station number US1INGR0024 on date 5/3/2002.
- TMAX outlier of 21.7 deg C found in station number US1INGR0024 on date 3/1/2014.
- TMAX outlier of 25 deg C found in station number US1INLG0001 on date 4/5/2007.
- TMIN outlier of -11.1 deg C found in station number US1INLG0005 on date 6/24/2003.
- TMIN outlier of -17.8 deg C found in station number US1INLG0005 on date 8/15/2003.
- TMIN outlier of -17.8 deg C found in station number US1INLG0005 on date 8/19/2003.
- TMIN outlier of -17.8 deg C found in station number US1INLG0005 on date 8/20/2003.
- TMIN outlier of -17.8 deg C found in station number US1INLG0005 on date 9/11/2003.
- TMIN outlier of -17.8 deg C found in station number US1INLG0005 on date 9/12/2003.
- TMIN outlier of -16.1 deg C found in station number US1INLG0005 on date 9/13/2003.
- TMIN outlier of -16.1 deg C found in station number US1INLG0005 on date 9/14/2003.
- TMIN outlier of -17.8 deg C found in station number US1INLG0005 on date 9/17/2003.
- TMIN outlier of -17.7 deg C found in station number US1INLG0006 on date 7/16/2006.
- TMAX outlier of -13.3 deg C found in station number US1INLG0013 on date 7/11/2000.
- TMAX outlier of -13.3 deg C found in station number US1INLG0013 on date 7/28/2000.
- TMIN outlier of -17.1 deg C found in station number US1INNB0004 on date 6/6/2010.
- TMAX outlier of 23.9 deg C found in station number US1INNB0006 on date 2/17/2007.
- TMIN outlier of -8.3 deg C found in station number US1INNB0024 on date 8/20/2001.
- TMIN outlier of -17.7 deg C found in station number US1INNB0024 on date 8/11/2007.
- TMIN outlier of -5.5 deg C found in station number US1INNB0024 on date 7/5/2012.

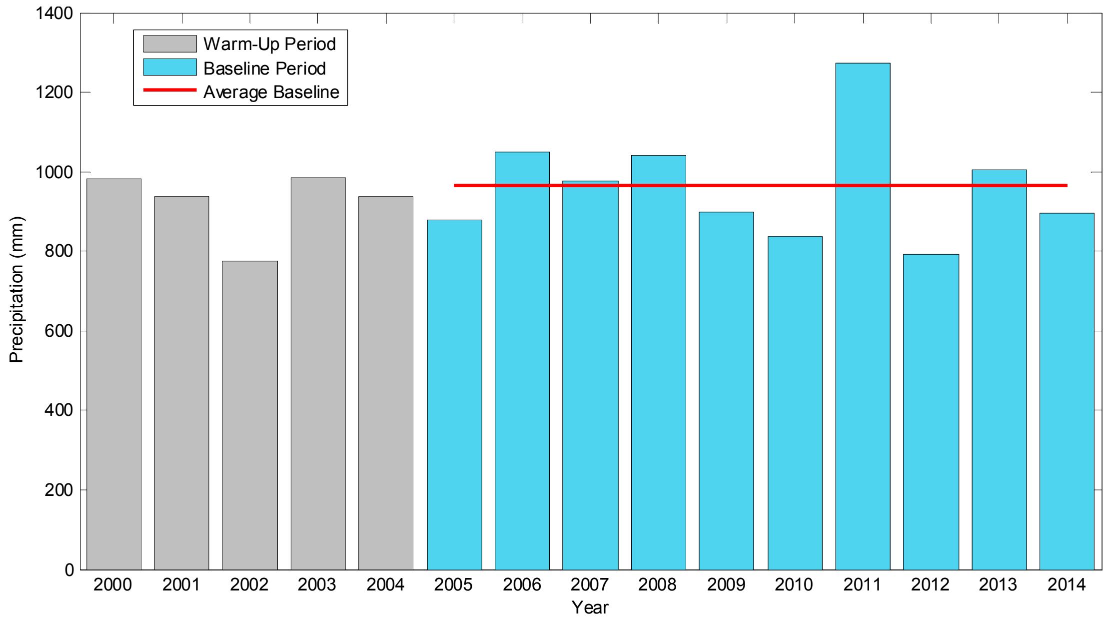


Figure A4.2: Average annual precipitation in the Maumee River Watershed stations.

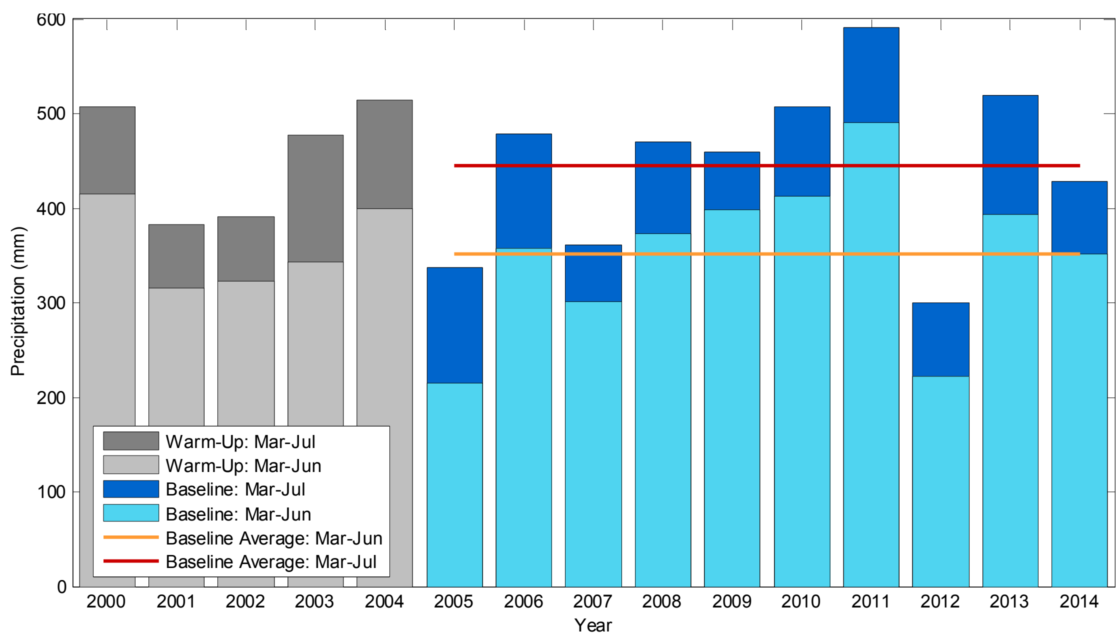


Figure A4.3: Average spring precipitation in the Maumee River Watershed using two different definitions of spring: (1) March - June, and (2) March - July. GLWQA targets are based on spring defined as March - July.

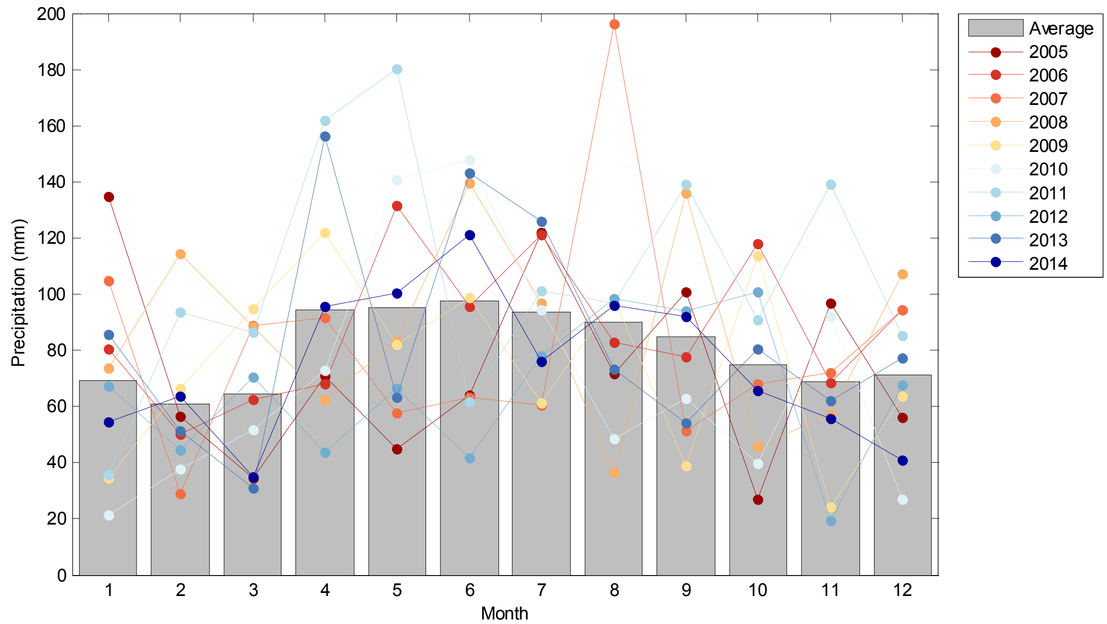


Figure A4.4: Average monthly precipitation during 2005 - 2014, showing variation among the years.

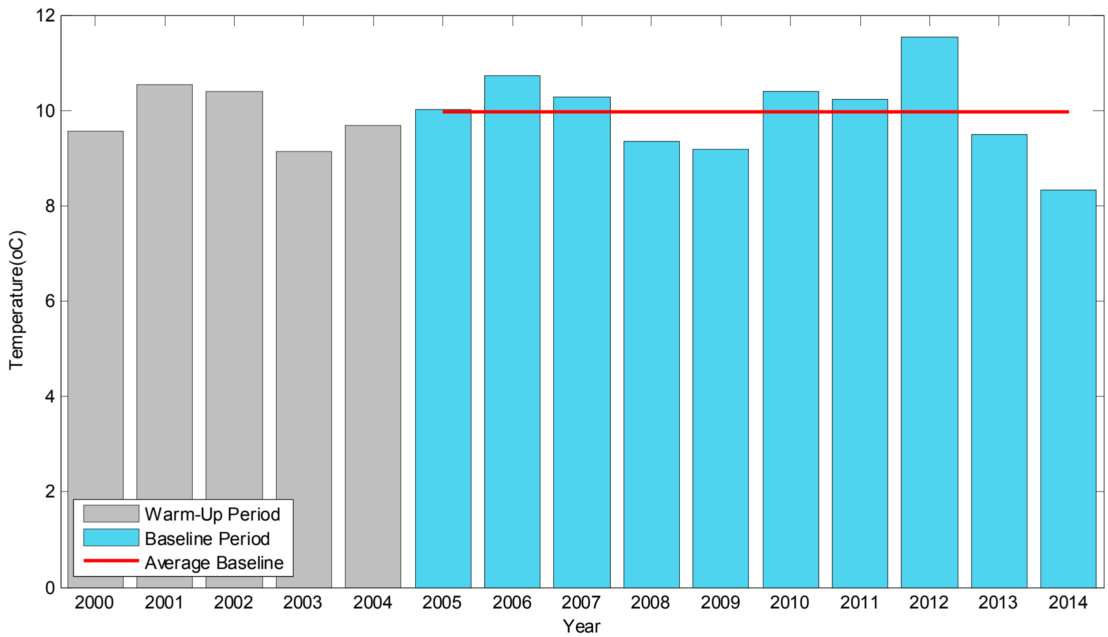


Figure A4.5: Annual average temperature across the Maumee River watershed from 2000 - 2015.

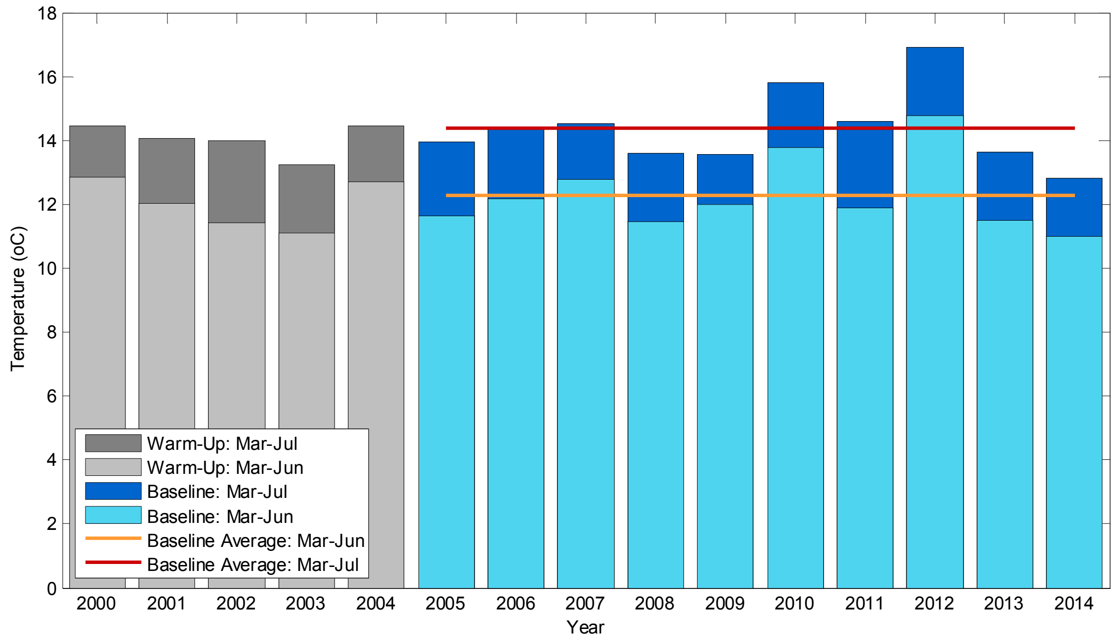


Figure A4.6: Average spring temperatures in the Maumee River Watershed using two different definitions of spring: (1) March - June, and (2) March - July. GLWQA targets are based on spring defined as March - July.

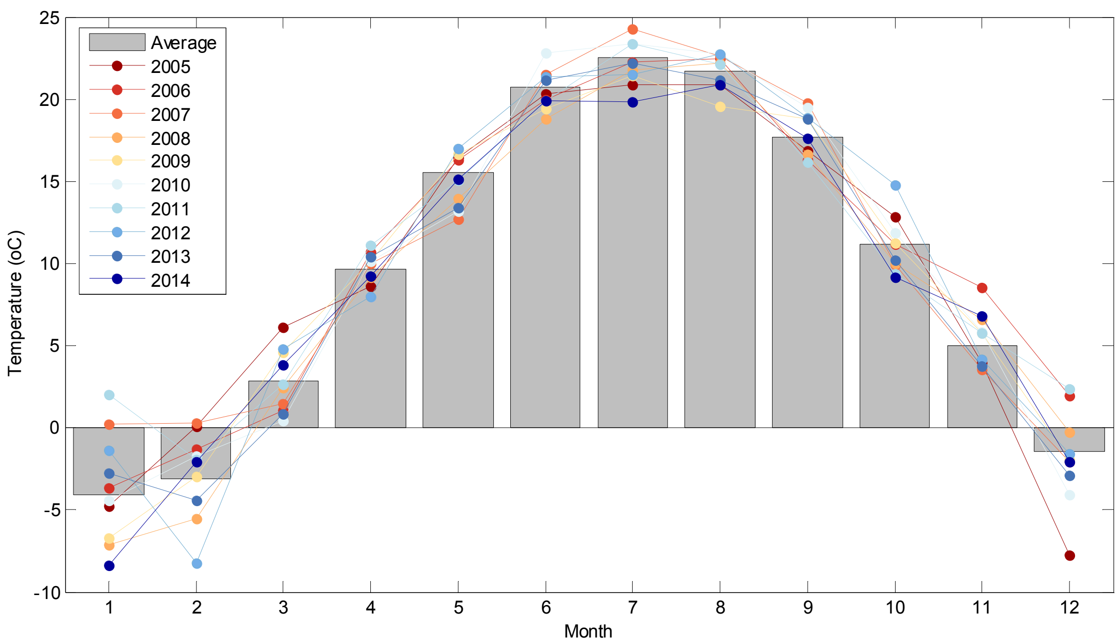


Figure A4.7: Average monthly temperature during 2005 - 2014, showing variation among the years.

References

Menne MJ, Durre I, Korzeniewski B, McNeal S, Thomas K, Yin X, Anthony S, Ray R, Vose RS, Gleason BE, Houston TG. 2012: Global Historical Climatology Network - Daily (GHCN-Daily), Version 3.21. NOAA National Climatic Data Center. http://doi.org/10.7289/V5D21VHZ [Accessed 2 August 2015]. Retrieved from ftp://ftp.ncdc.noaa.gov/pub/data/ghcn/daily/.

A5. Details on baseline point source data

Point source discharge data for common Baseline models were retrieved from the EPA's Discharge Monitoring Report (DMR) Pollutant Loading Tool, available at http://cfpub.epa.gov/dmr/ez_search.cfm, which is based on data submitted by the National Pollutant Discharge Elimination System (NPDES) permit holders. All stations falling within the Maumee River watershed were retrieved, and a point source shapefile was created to map station locations (Figure A5.1). The locations of major and minor point sources are shown in Figure A5.1.

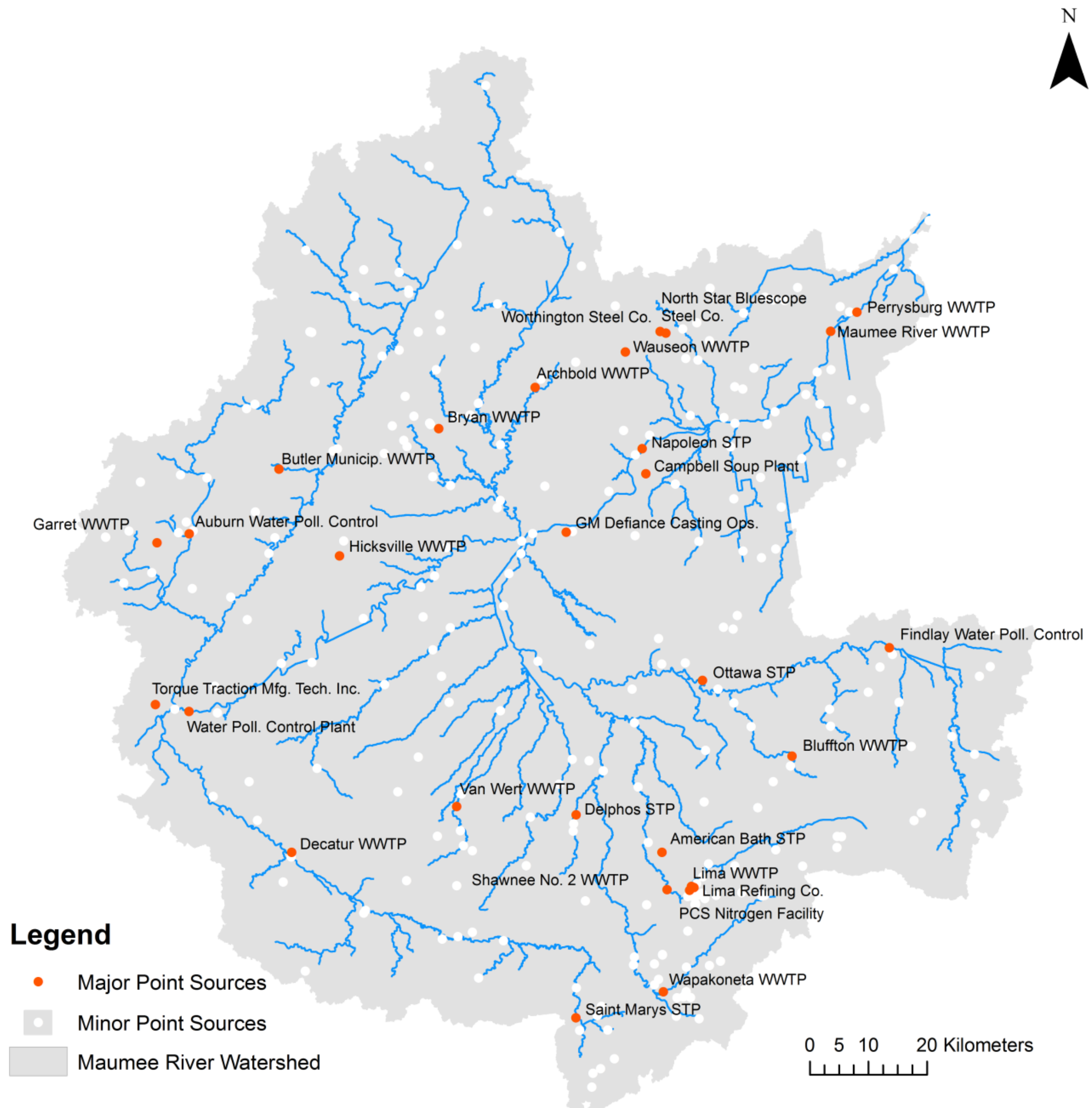


Figure A5.1: Location of all point sources in the Maumee River Watershed. Major point sources highlighted in red and identified.

Data for each station from October 2008 (the earliest data available on the DMR website) to June 2015 were summarized into monthly averages of flow, nutrient, and sediments to be added to SWAT and monthly averages were used to create annual point source inputs for SPARROW. Detailed information on the data collected from the DMR system and how it relates to SWAT are below:

- Data collected from DMR:
  1. The *FLOMON* field in SWAT requires the average daily water discharge for a month. Data for this field were usually from the DMR system parameter: *Flow (in conduit or through treatment plant)* (parameter code: *50050*). A few facilities instead reported *Flow Rate* (parameter code: *00056*) for this measurement.
  2. The *SEDMON* field in SWAT requires the average daily sediment loading for a given month. Data for this field were from the DMR system parameter: Total Suspended Solids (parameter code: *00530*).
  3. The *NO3MON* field in SWAT requires the average daily nitrate loading for a given month. Data for this field were usually from the DMR system parameter: *Total Nitrate and Nitrite* (parameter code: *00630*). A few facilities instead reported *Total Nitrate* (parameter code: *00620*) for this measurement.
  4. The *NH3MON* field in SWAT requires the average daily ammonia loading for a given month. Data for this field were from the DMR system parameter: *Total Ammonia* (parameter code: *00610*).
  5. The *ORGNMON* field in SWAT requires the average daily organic nitrogen loading for a given month. Data for this field were calculated by subtracting DMR system parameter *NH3mon* from *Total Kjeldahl Nitrogen* (parameter code: *00625*). If the result was negative, it was assumed that there were no organic nitrogen contributions and therefore the value should be zero.
  6. The *ORGPMON* field in SWAT requires the average daily organic phosphorus loading for a given month. Data for this field were calculated from the DMR system parameter: *Total Phosphorus* (parameter code: *00665*) multiplied by 0.53 (Bosch, personal communication, July 16, 2015).
  7. The *MINPMON* field in SWAT requires the average daily mineral (soluble) phosphorus loading for a given month. Data for this field are calculated from the DMR system parameter: *Total Phosphorus* (parameter code: *00665*) multiplied by 0.47 (Bosch, personal communication, July 16, 2015).
- Unit conversions and formatting for SWAT:
  1. A facility may report the measurements in *quantity* or *concentration*. If the reporting measurement is *Quantity1*, unit conversion was performed based on the required unit of the RECMON.DAT file. For example, DMR usually reports the flow in the unit of *Million Gallons per day* (MGD). These values were converted to cubic meter per day for the RECMON.DAT file by multiplying by the value 3,785.4. DMR usually

reports the total suspended solids in the unit of *kilogram per day*. These values were converted to *tons per day* for the RECMON.DAT file by multiplying by the value 0.001.

2. If the reporting measurement was *Concentration2*, the values were converted to loads by multiplying the concentration by the flow. The concentration measurements were usually in the unit of *milligrams per liter*, thus the conversion to *kilograms per day* is based on this equation: *Concentration* (*mg/L*) × *flow* (*MGD*) × 3.7854 = *Quantity* (*kg/day*).
3. Some facilities had more than one outfall. If so, the loads and flows from all the outfalls were added to represent the total amount for that facility.
4. The monthly average was calculated by averaging all occurrences of each month at a facility.
    a. If a column contained only missing values, it is likely that DMR does not provide any records for this parameter. Since SWAT does not allow for a "NaN" value, these missing values were replaced with zeros.
    b. If a column contains values for some time periods, but not others, it was assumed that the NPDES data either wasn't reported for some unknown reason or that the NPDES permit does not require monthly reporting for that constituent. As a result, these missing values were replaced with the mean of the values from the same column, which was the average monthly value.

- Removing outliers
    1. After processing the data as outlined above, the data was further examined for outliers. Here we defined an outlier as a value that is 250 times larger than or equal to the median of all the non-zero monthly flow values. We first calculated the median of all the non-zero monthly flow values of each facility. We then filtered all the monthly values of that facility to see if there are any outliers. Eventually, among all the monthly flow values, 15 values from 10 different facilities in Indiana and Ohio were determined to be outliers. The data managers of EPA's DMR confirmed that these outliers resulted from decimal errors (misplaced the decimal points of the number) or reporting errors (performed incorrect unit conversion) (Jeff Ewick and Eddie Swindall, personal communication on August 13, 2015 and August 17, 2015, respectively). As a consequence, the flow outliers and the resultant sediment and nutrient monthly values were removed from the dataset and filled in by calculating the average value for that month at that facility. When removing the flow outliers, all sediment and nutrient outliers were also removed. Figure A5.2 shows the Maumee Watershed phosphorus point source loads with outliers included, and Figure A5.3 shows the data with outliers removed to demonstrate the impact of these few outliers.

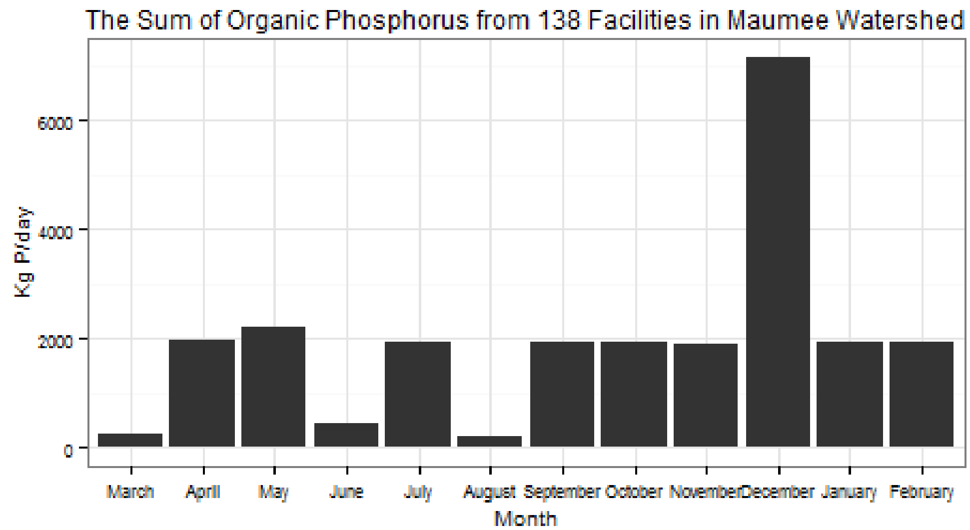


Figure A5.2: Point source phosphorus contributions summed up over the Maumee River watershed, including outlier data points.

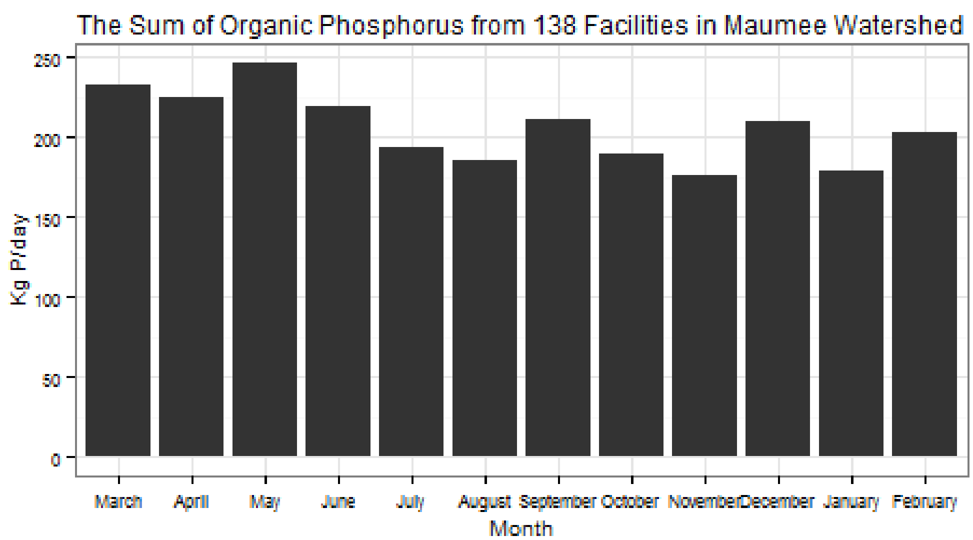


Figure A5.3: Point source phosphorus contributions summed up over the Maumee River watershed, removing outlier data points.

2. Due to the nature of the code, and the fact that it fills in missing data with average monthly values, the two facilities with abnormally high flow values were not only impacting the average values for the months they occurred, they were also impacting the values on the months where data was missing and averages were used.

The final point source data were summarized across the entire Maumee River Watershed, and results are provided in Figure A5.4-5.7. The relative contributions of individual stations are shown in Figures A5.8 and A5.9; an important note is that only including facilities listed as ‘major’ by the EPA could lead to under prediction, especially for phosphorus, as this category is typically defined by total discharge alone.

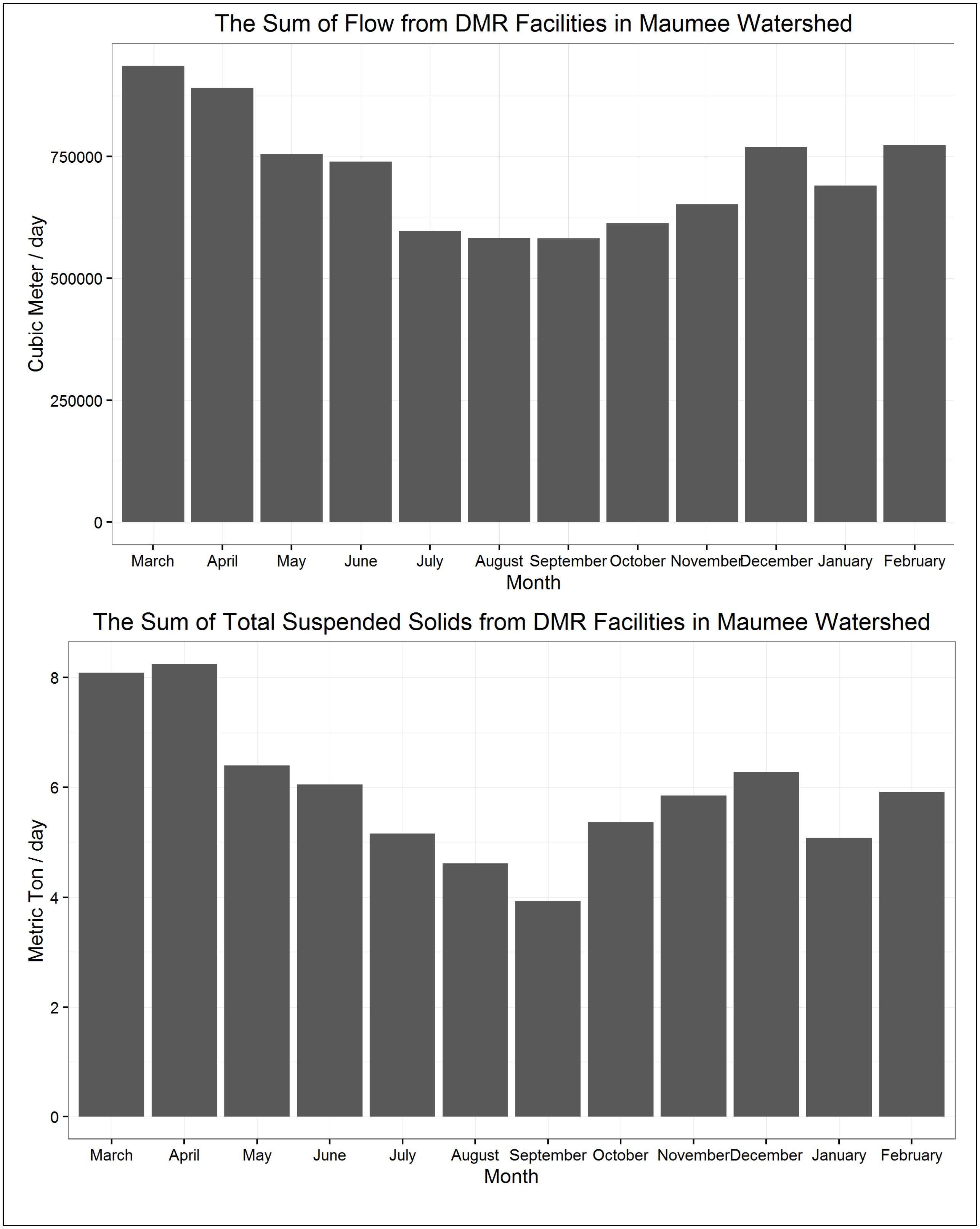


Figure A5.4: Daily point source contributions summed over the entire Maumee River Watershed for flow (top) and total suspended solids (bottom)

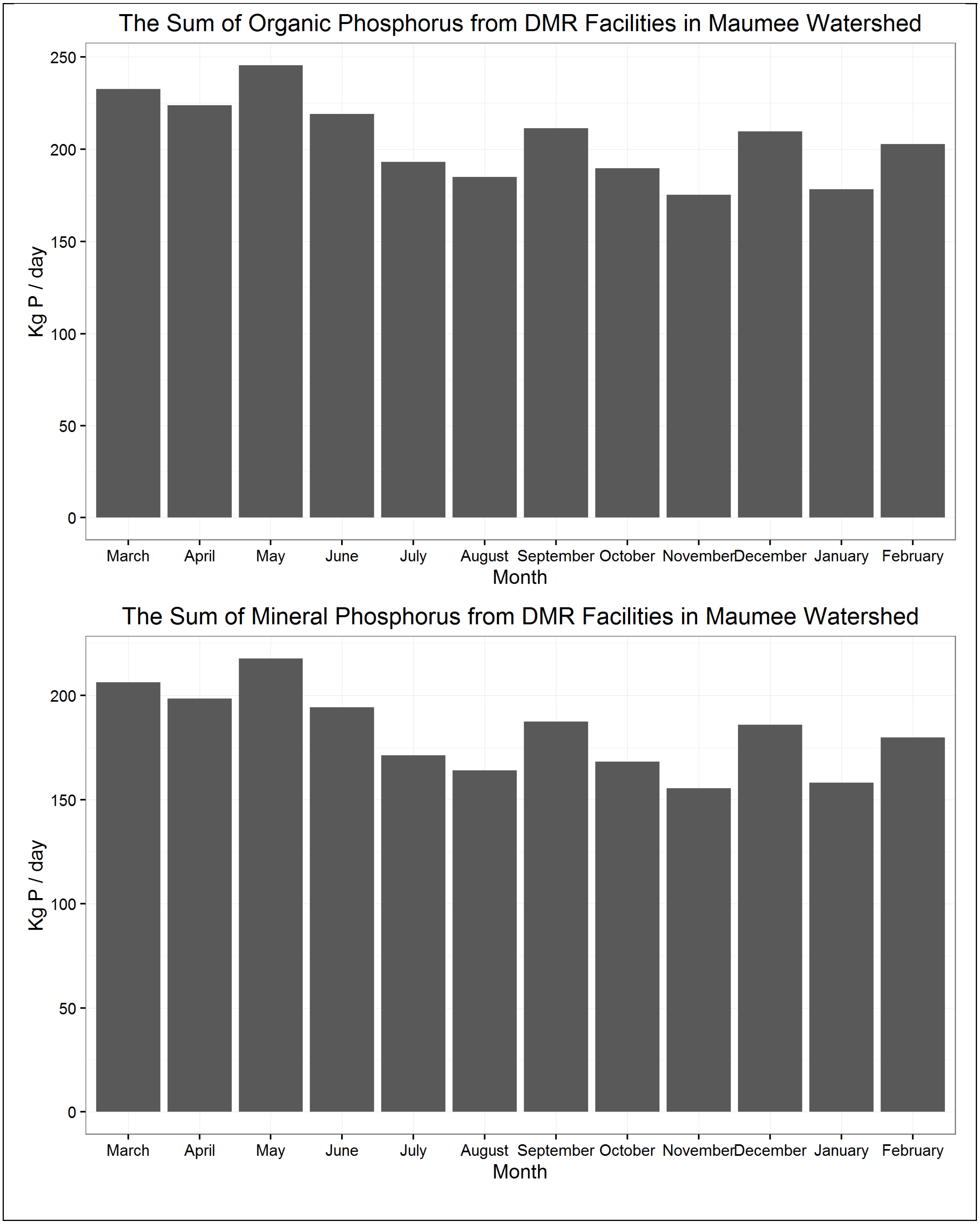


Figure A5.5: Daily point source contributions summed over the entire Maumee River Watershed for organic P (top) and mineral P (bottom)

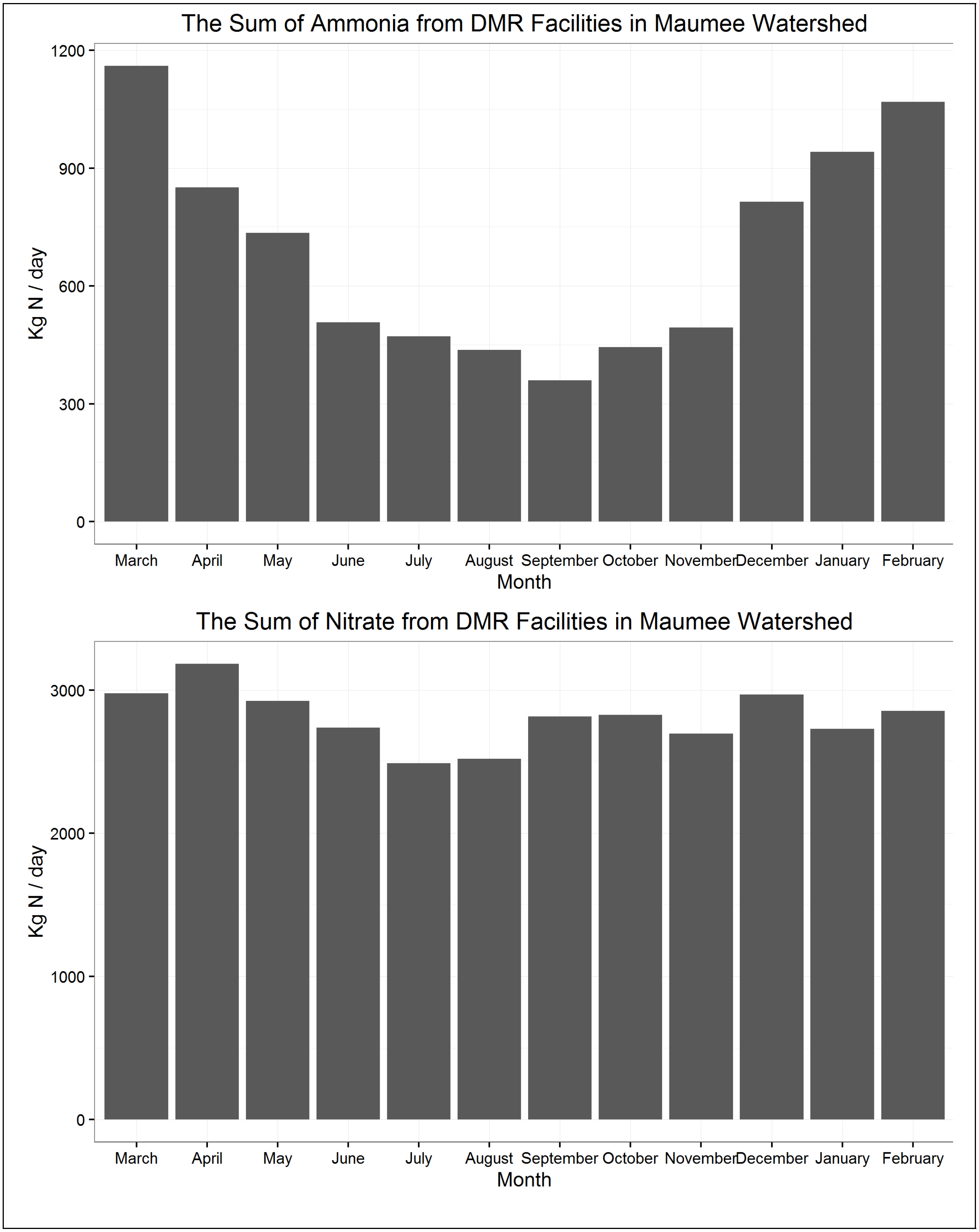


Figure A5.6: Daily point source contributions summed over the entire Maumee River Watershed for ammonia (top) and nitrate (bottom)

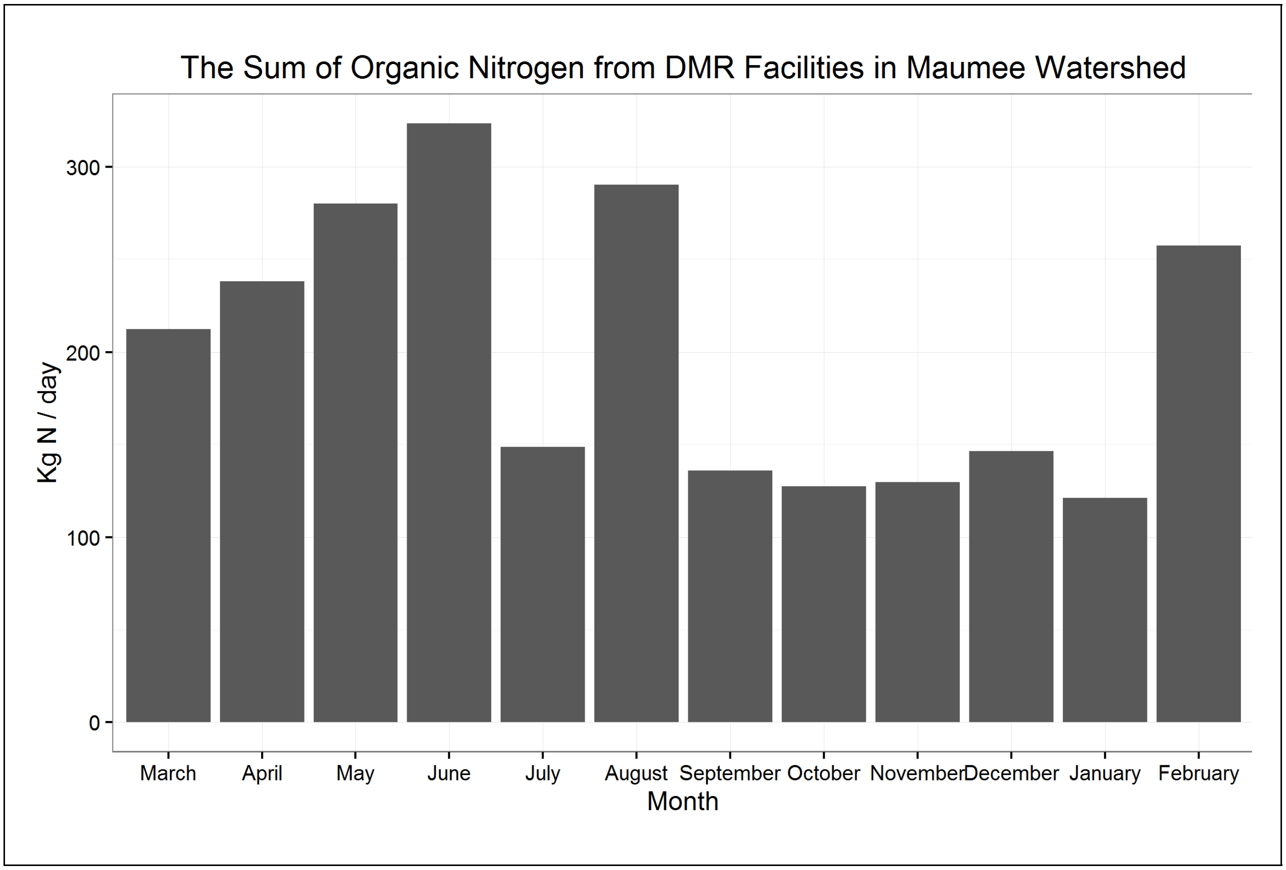
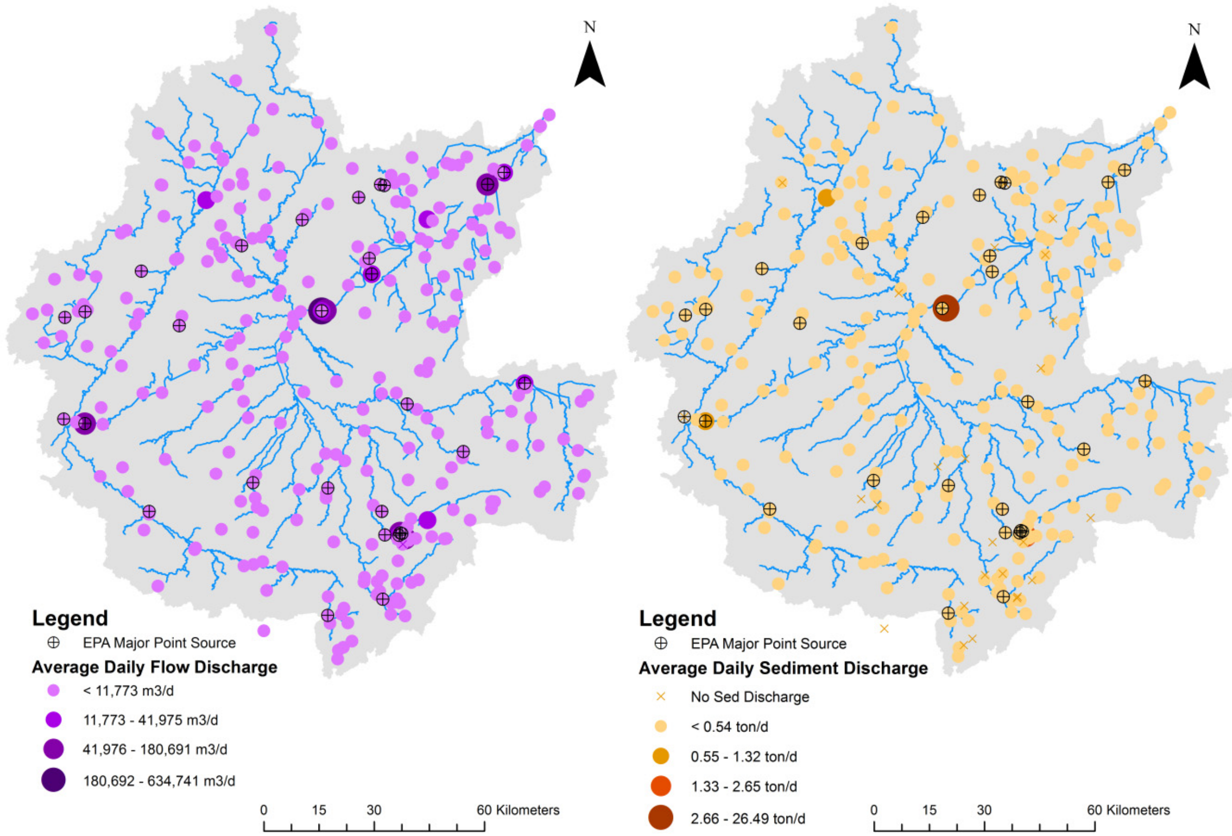


Figure A5.7: Daily point source contributions summed over the entire Maumee River Watershed for organic nitrogen



Average daily flow discharged from data available through the EPA DMR tool from 2007-2015. Compiled in August of 2015.

Average daily sediment discharge from data available through the EPA DMR tool from 2007-2015. Compiled in August of 2015.

Figure A5.8: Daily flow discharge and sediment discharge from Maumee Watershed point sources by facility. Sources contributing more flow or sediment are identified with larger and darker circles.

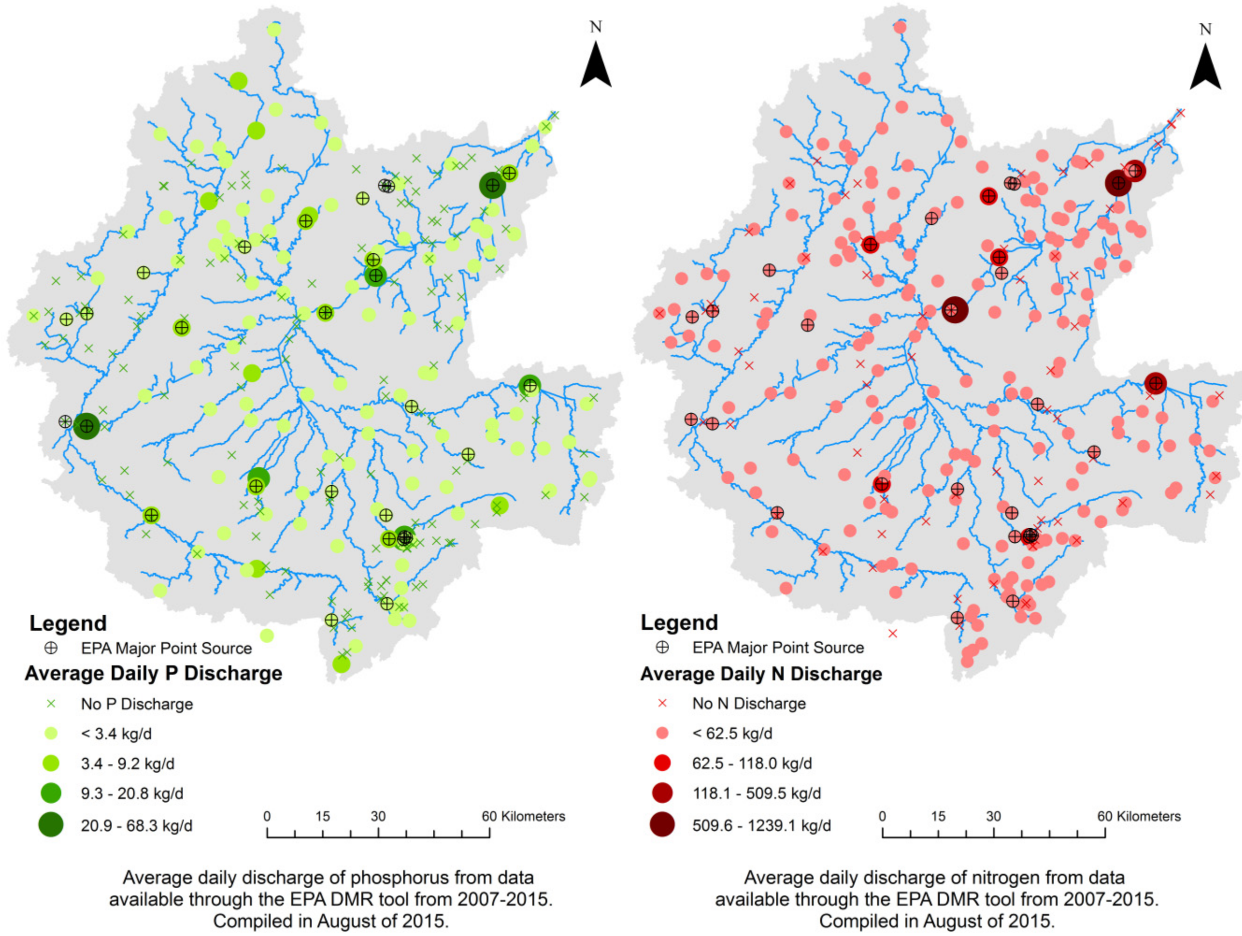


Figure A5.9: Daily phosphorus discharge (left) and nitrogen discharge (right) from Maumee Watershed point sources by station. Sources contributing more phosphorus or nitrogen are identified with larger and darker circles.

References

Bosch, N. July 16, 2015. Personal communication regarding the ratio of mineral P to organic P based on discharge measurements from the Toledo, OH wastewater treatment plant.

A6. Details on observed data used for validating the models

Streamflow data were gathered from the United States Geological Survey National Water Information System for the Maumee River at Waterville, OH gage (#04193500): http://waterdata.usgs.gov/nwis/dv/?site_no=04193500&agency_cd=USGS&referred_module=sw. Measured water quality data used to calibrate the SWAT models were downloaded from the National Center for Water Quality Research (NCWQR) at Heidelberg University available at: http://www.heidelberg.edu/academiclife/distinctive/ncwqr/data/data. The Waterville station in the Maumee River watershed was used to calibrate and validate all the models. Loads were derived by converting concentrations reported by Heidelberg University data to loads using USGS streamflow data. Missing data were filled in using a method specifically designed for the Maumee River at Waterville, Ohio (Obenour et al. 2014). Missing TP data were filled in by calculating daily concentration using a linear regression of concentration on daily flow for the 20 days closest to the missing data point. Missing data for all other constituents were filled in using an average concentration in the nearest 10 days. Table A6.1 describes the amount of days missing and how much of the load the missing days accounted for. This complete dataset was used to make plots for outputs. In order to calculate baseline statistics, however, if a month had more than 14 days of no data, that month was removed to better assess model performance (Table A6.2) Observed values were plotted on a monthly, annual, and spring (March - July) basis, shown in Figures A6.1-A6.10.

Table A6.1: Number of missing days for each water quality constituent and the percent of estimated load compared with total load during the time period.

Constituent	% of days	% of estimated entire load
Sediment	6%	9%
TP	5%	11%
DRP	7%	9%
TN	12%	10%
Nitrate	7%	10%

Table A6.2: Months of data removed for calculation of validation statistics for each constituent due to the total amount of missing data for that month being greater than 14 days.

Constituent	Year	Month	#NaNs	Constituent	Year	Month	#NaNs
Sediment	2005	January	22	TN	2005	January	22
	2006	January	21		2005	July	16
	2011	July	20		2006	January	21
TP	2005	January	22		2007	July	18
	2006	January	21		2011	July	20
	2011	July	20		2013	September	28
DRP	2005	January	22		2014	February	24
	2005	June	15	Nitrate	2005	January	22
	2006	January	21		2006	January	21
	2011	July	20		2011	July	20
					2014	February	24

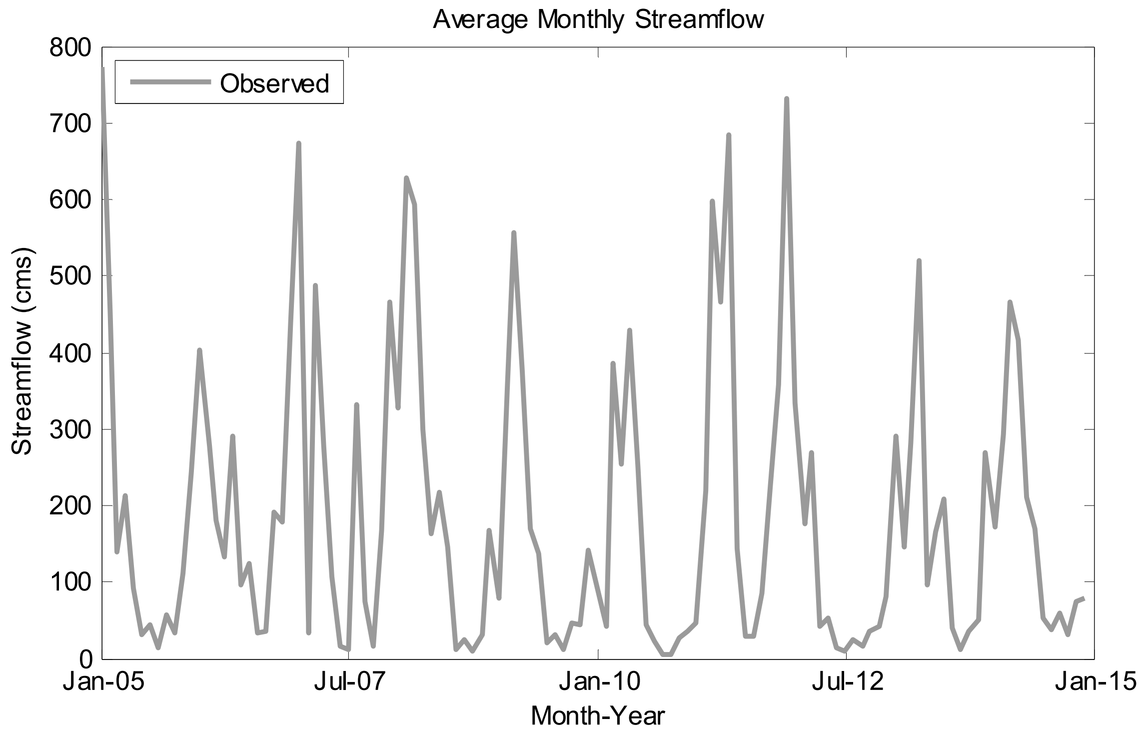


Figure A6.1: Average monthly observed flows from 2005-2014.

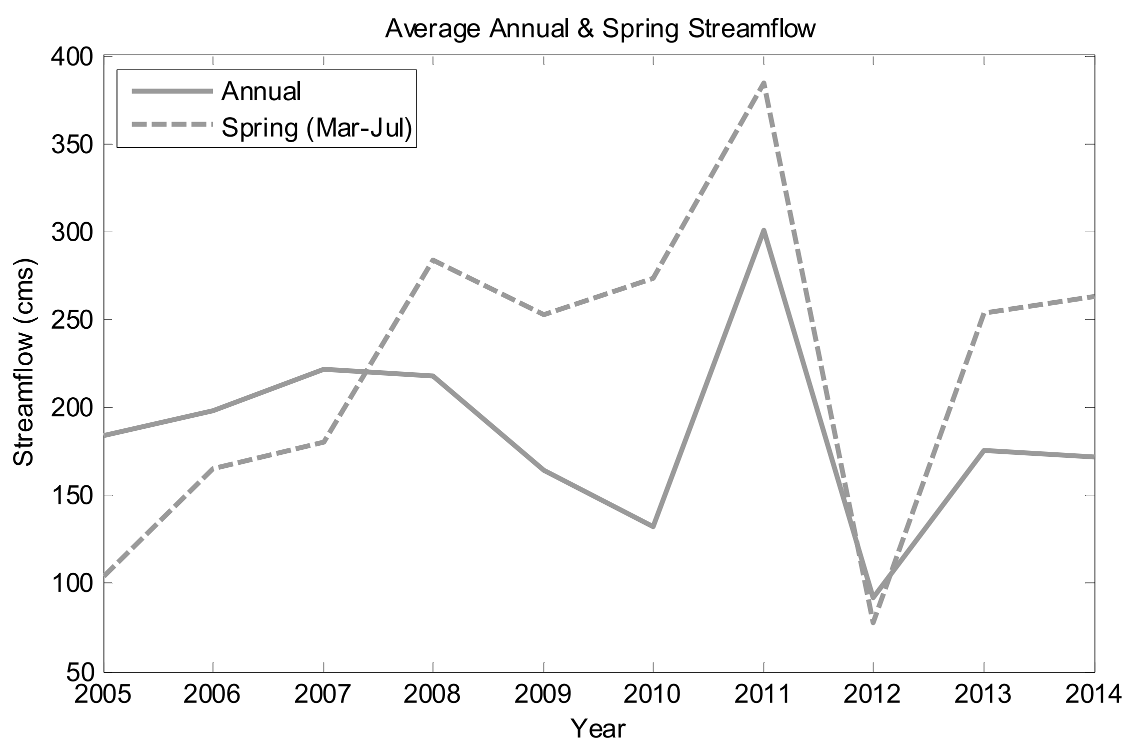


Figure A6.2: Average annual and average spring (March-July) streamflow from 2005-2014.

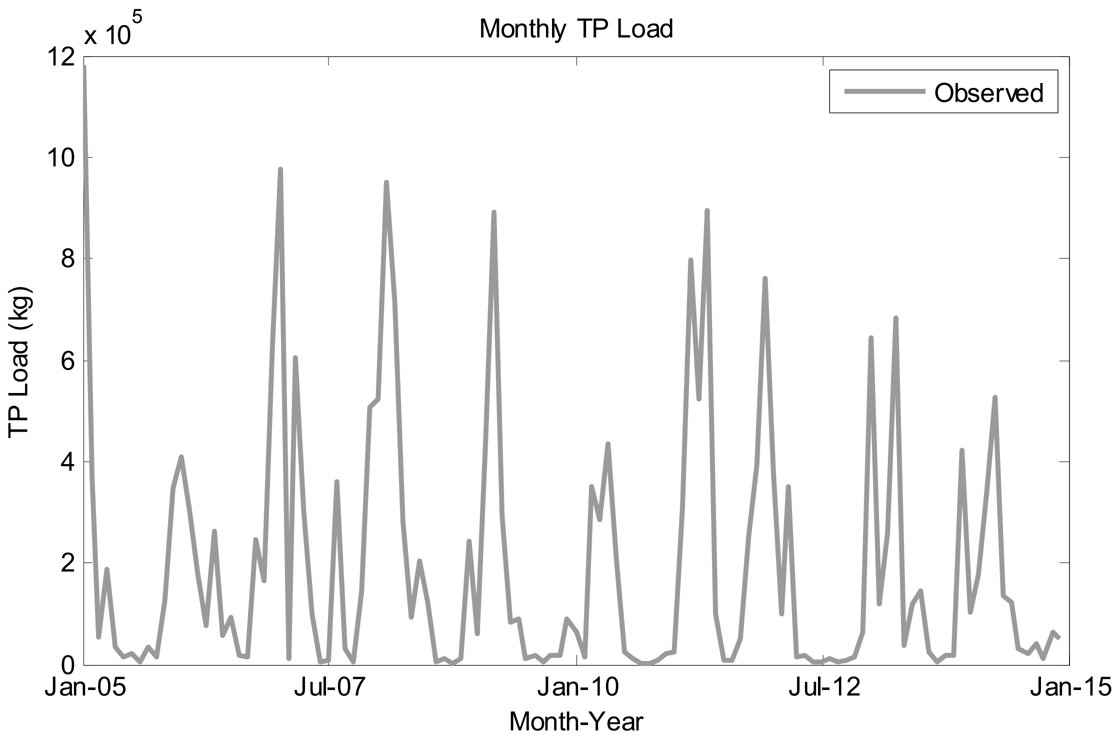


Figure A6.3: Monthly total phosphorus loads from 2005-2014.

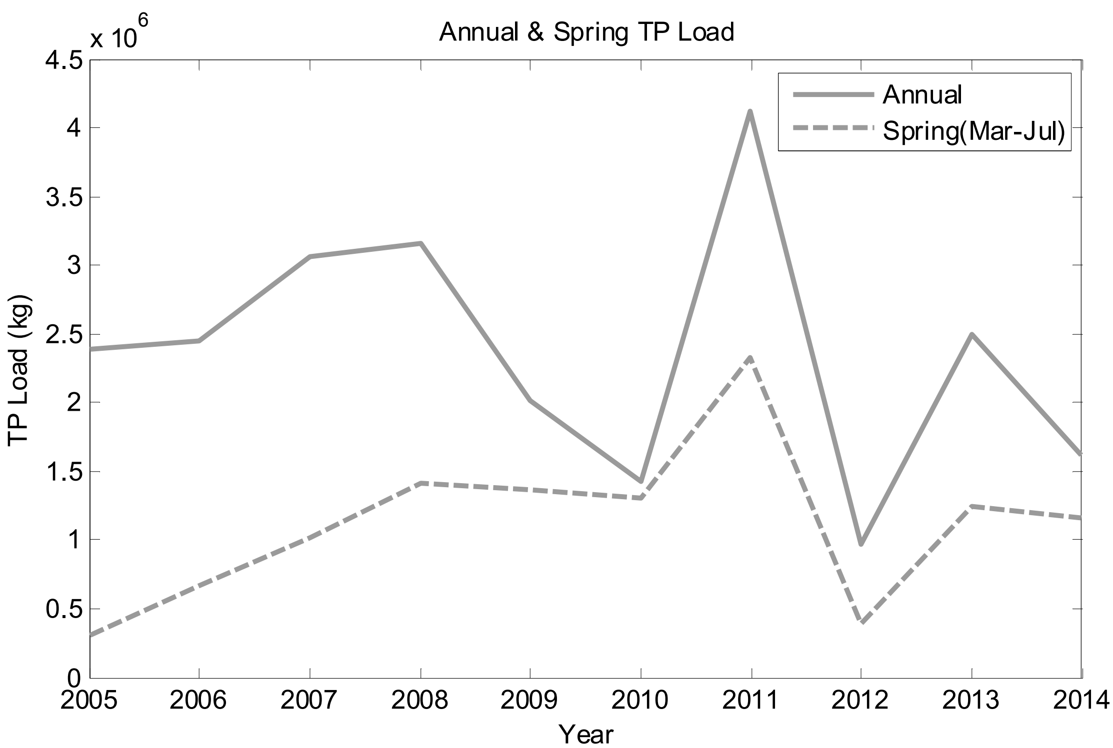


Figure A6.4: Annual and spring (March-July) total phosphorus loads for 2005-2014.

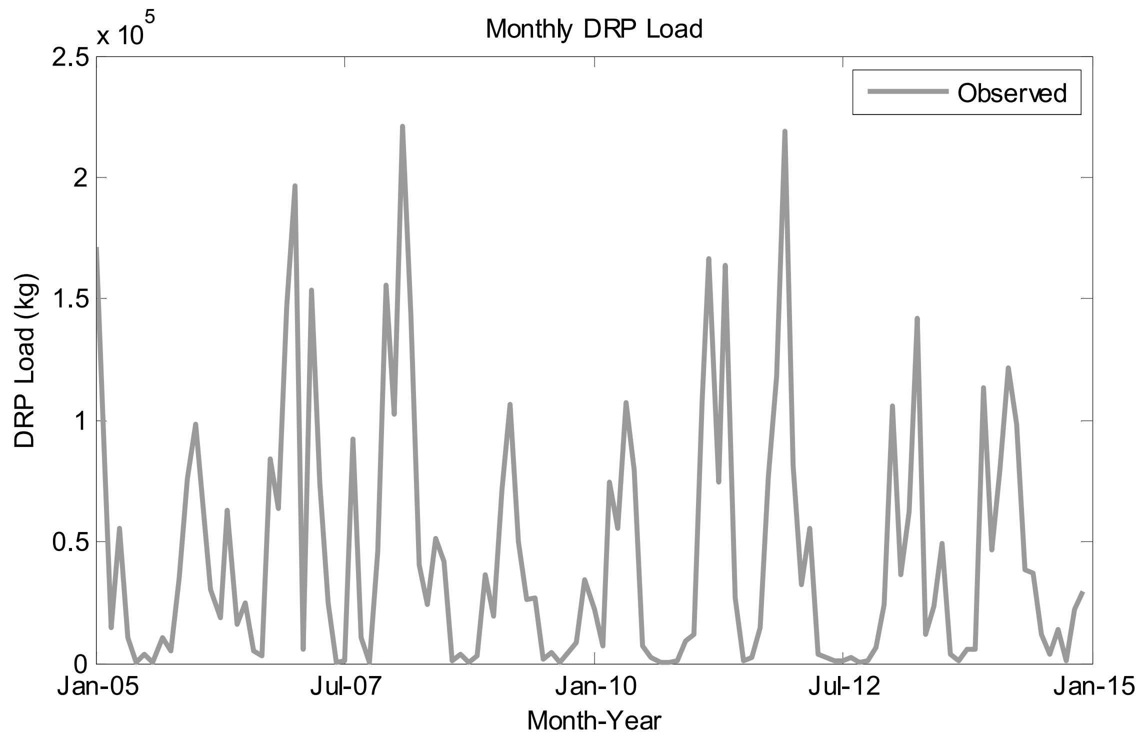


Figure A6.5: Monthly dissolved reactive phosphorus loads for 2005 - 2014.

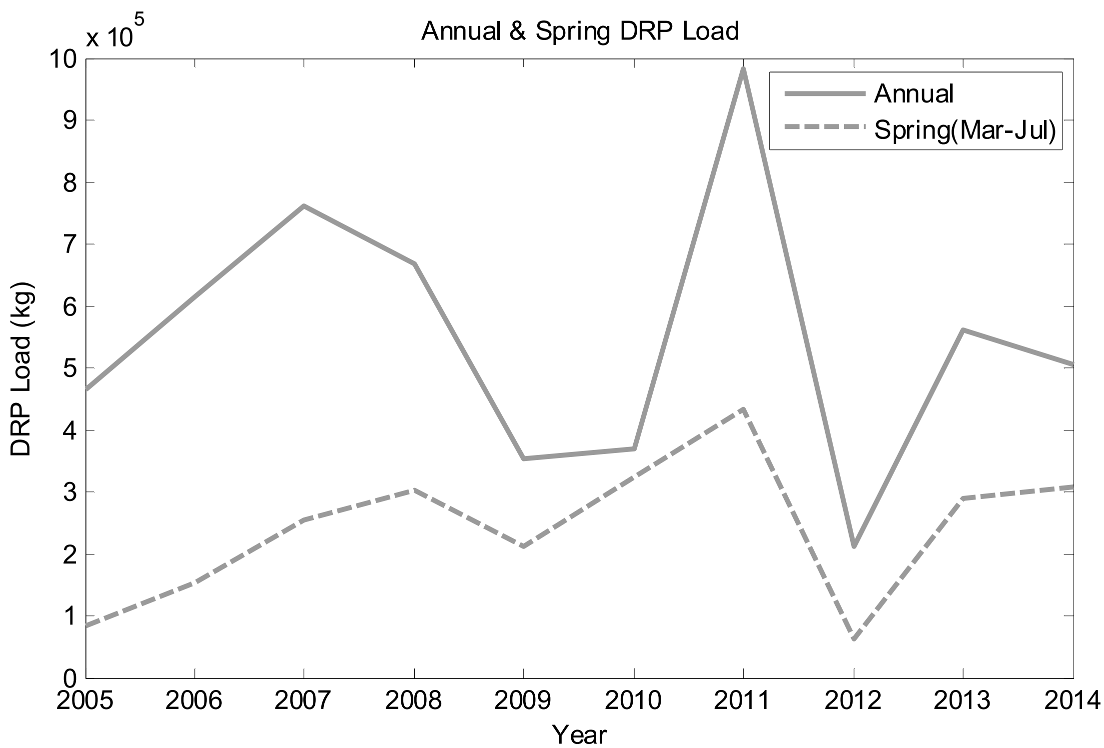


Figure A6.6: Annual and spring (March - July) dissolved reactive phosphorus loads from 2005 - 2014.

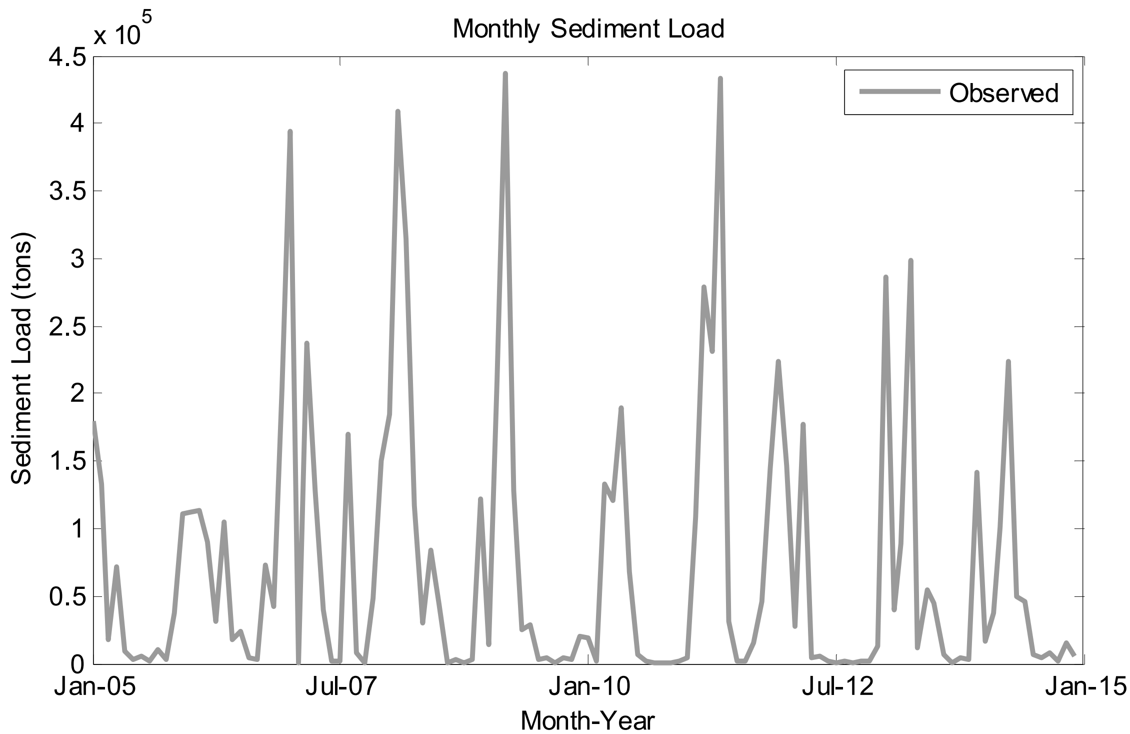


Figure A6.7: Monthly sediment load from 2005 - 2014.

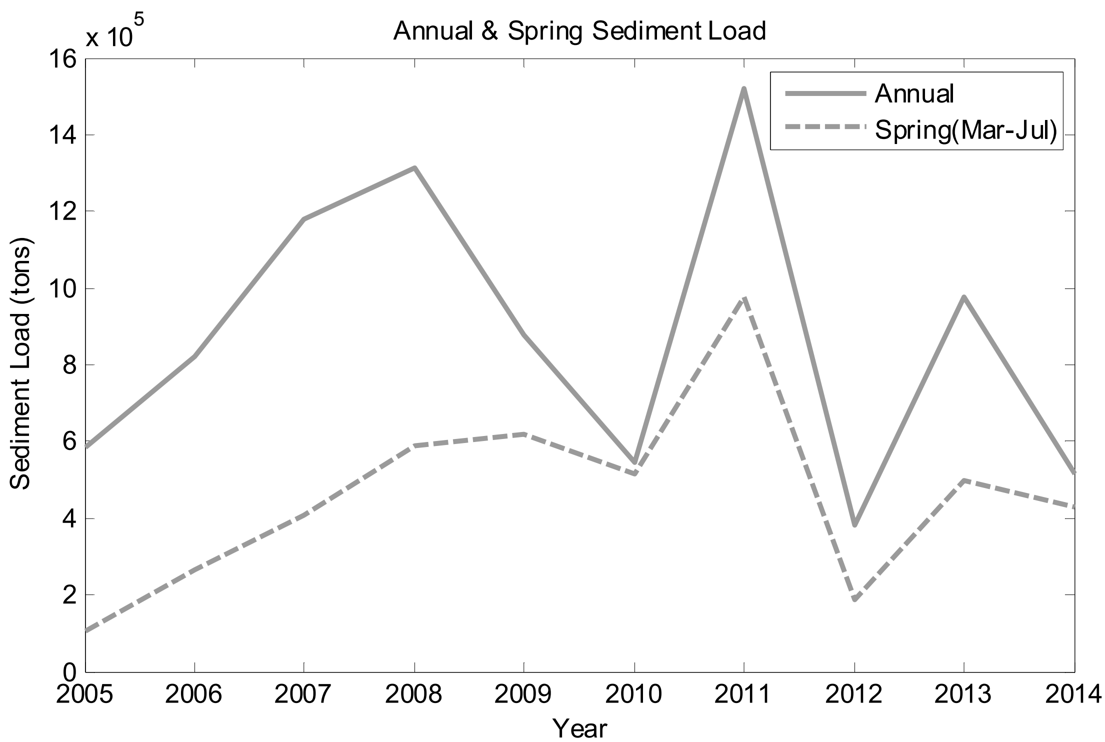


Figure A6.8: Annual and spring (March - July) sediment loads from 2005 - 2014.

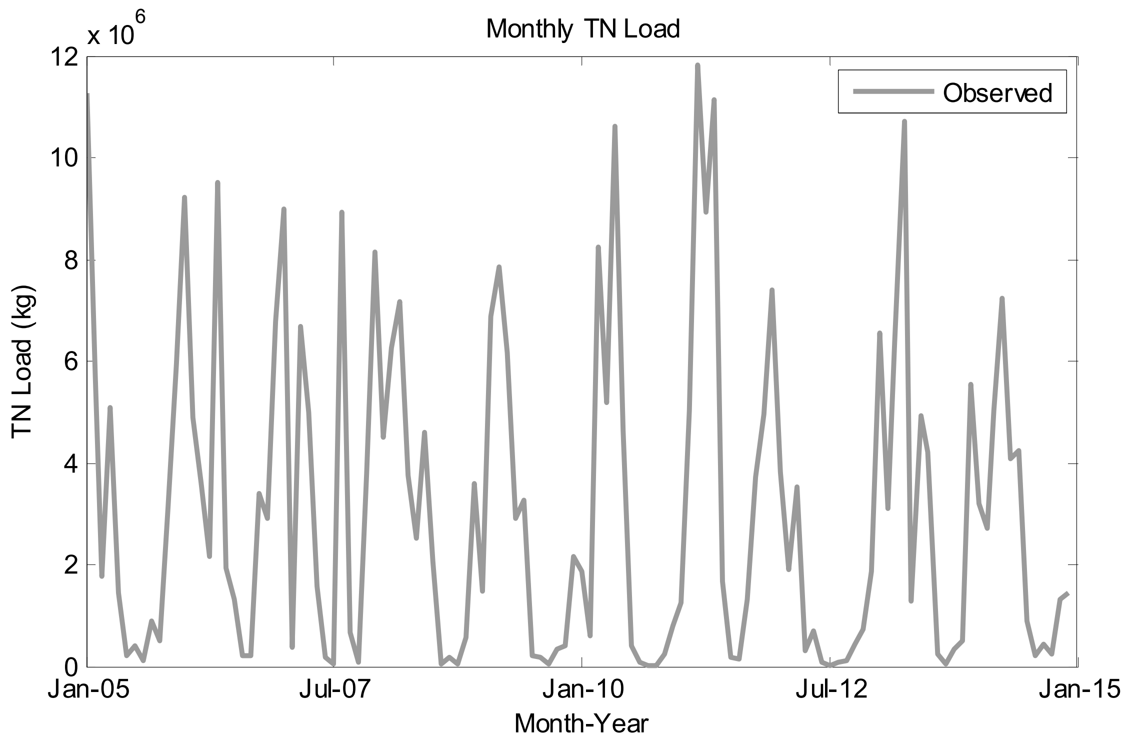


Figure A6.9: Monthly total nitrogen load from 2005 - 2014.

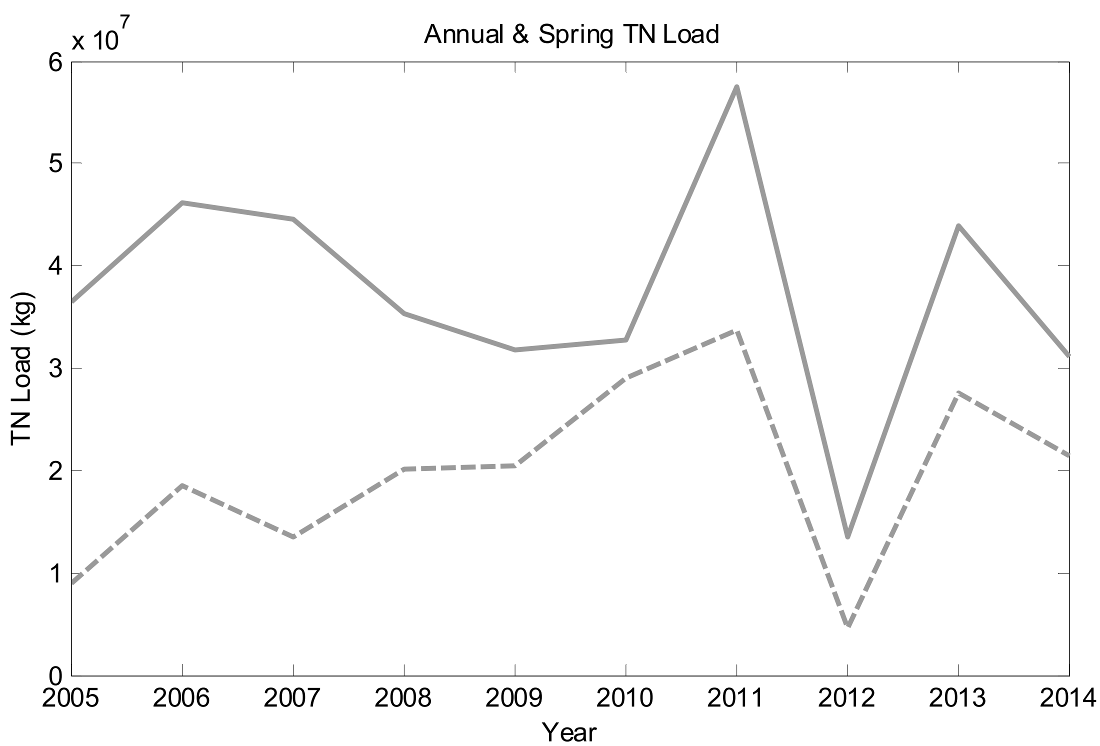


Figure A6.10: Annual and spring (March - July) total nitrogen loads for 2005 - 2014.

References

Obenour DR, Groneworld AD, Stow CA, and Scavia D. 2014. Using a Bayesian hierarchical model to improve Lake Erie cyanobacteria bloom forecasts. Water Resources Research, 50: 7847-7860, available online at: http://graham.umich.edu/scavia/wp-content/uploads/2009/11/Obenour-et-al-2014-WRR.pdf.

A7. Single-practice scenario results

Single-practice scenarios were run to assess the differences among models as well as to help inform bundled scenario development. Table A7.1 describes the single-practice scenarios, along with details used in modeling the practices. All single-practice scenarios were simulated with the 5 SWAT models; the SPARROW model was only used to simulate the change in fertilizer rates and no point source scenarios. March-July and annual TP and DRP loads were extracted for each scenario and the results are provided in Figures A7.1-A7.4. It should be noted that the results for drainage water management scenarios are somewhat uncertain as the module has not been fully developed in SWAT and field studies showing the impact of DWM on P are not common, particularly with respect to changes in P loading in surface runoff.

Table A7.1: Single-practice scenarios, including modeling details. All scenarios were simulated in SWAT; *indicates scenario was simulated by SPARROW as well. †Equivalent to bundled scenario #1.

No.	Scenario	Description
1	Fertilizer placement: Subsurface fertilizer application	All cropland had 99% of fertilizer applications to the soil subsurface, using SWAT parameter FRT_SURFACE = 0.01 (1%). Tillage and fertilizer rates and timings remained the same as in the Baseline.
2*	Fertilizer rate: P fertilizer cut 50%	All cropland had P fertilizer rates reduced by 50% of the Baseline rates. N fertilizer rates were the same as in the Baseline.
3	Fertilizer timing I: P applied in spring	All cropland had P applied in the spring prior to planting corn and soybeans, rather than in the fall.
4	Fertilizer timing II: P applied in fall	All cropland had P applied in the fall, following harvest, and prior to planting corn and soybeans.
5	Cover crop: Cereal rye	All cropland had a cereal rye cover crop applied in all winters that the ground was bare in the Baseline.
6	Drainage water management: Testing the approach	All cropland had tile drains held near the soil surface (150 mm) over winter and summer months. Drainage water management was implemented by changing the depth of drain (DDRAIN) in the operations (.ops) files. For example, controls may be raised on 6/1, lowered on 10/1, raised again 10/ 31, and then lowered on 4/1.
7	Tillage: Continuous no-till	All cropland was managed without any tillage operations except for a no-till drill at crop planting. Fertilizer application, including placement, remained the same as in the Baseline, but no incorporation with tillage took place.
8	Crop rotation I: Continuous corn	All cropland was converted to a continuous corn crop rotation, using Baseline fertilizer and tillage methods for corn.
9	Crop rotation II: Winter wheat	All cropland was converted to a rotation including at least one year of wheat, using a Baseline wheat rotation in each of the models.
10	Buffer strips: High effectiveness	All cropland was given a buffer strip of high effectiveness, meaning a field area to buffer strip ratio of 22, a fraction of HRU draining to filter of 0.25, and a fraction of concentrated flow of 0.
11	Wetlands: Testing the approach	One 100-acre wetland was placed in all sub-basins greater than 100 acres in size; the wetland intercepted 100% of flow from the sub-basin.
12*†	No Point Sources: A theoretical test	All point source discharges were set to zero.

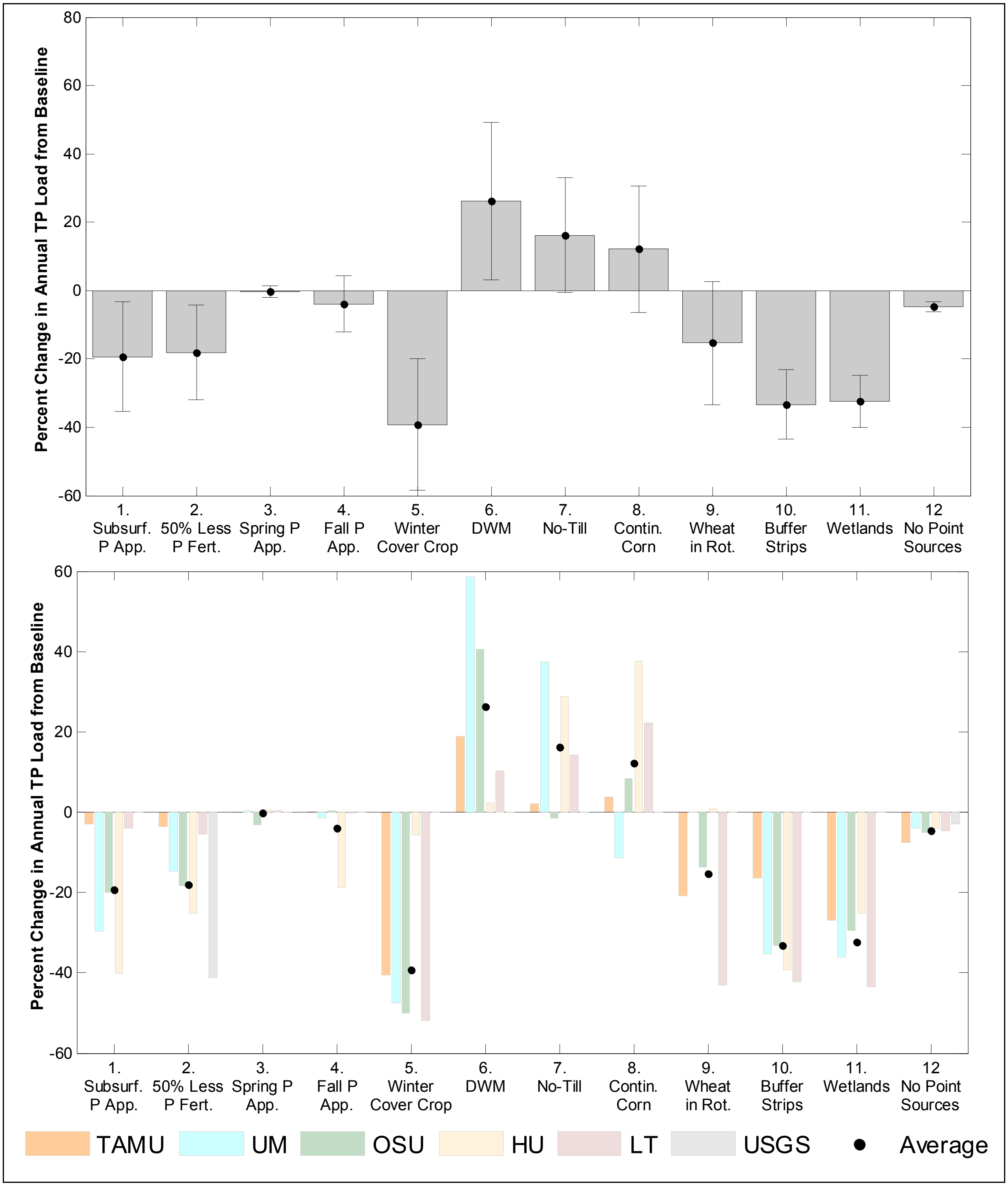


Figure A7.1: Percent change in annual TP loads from Baseline averaged across all models with standard deviations (top) and for each model (bottom).

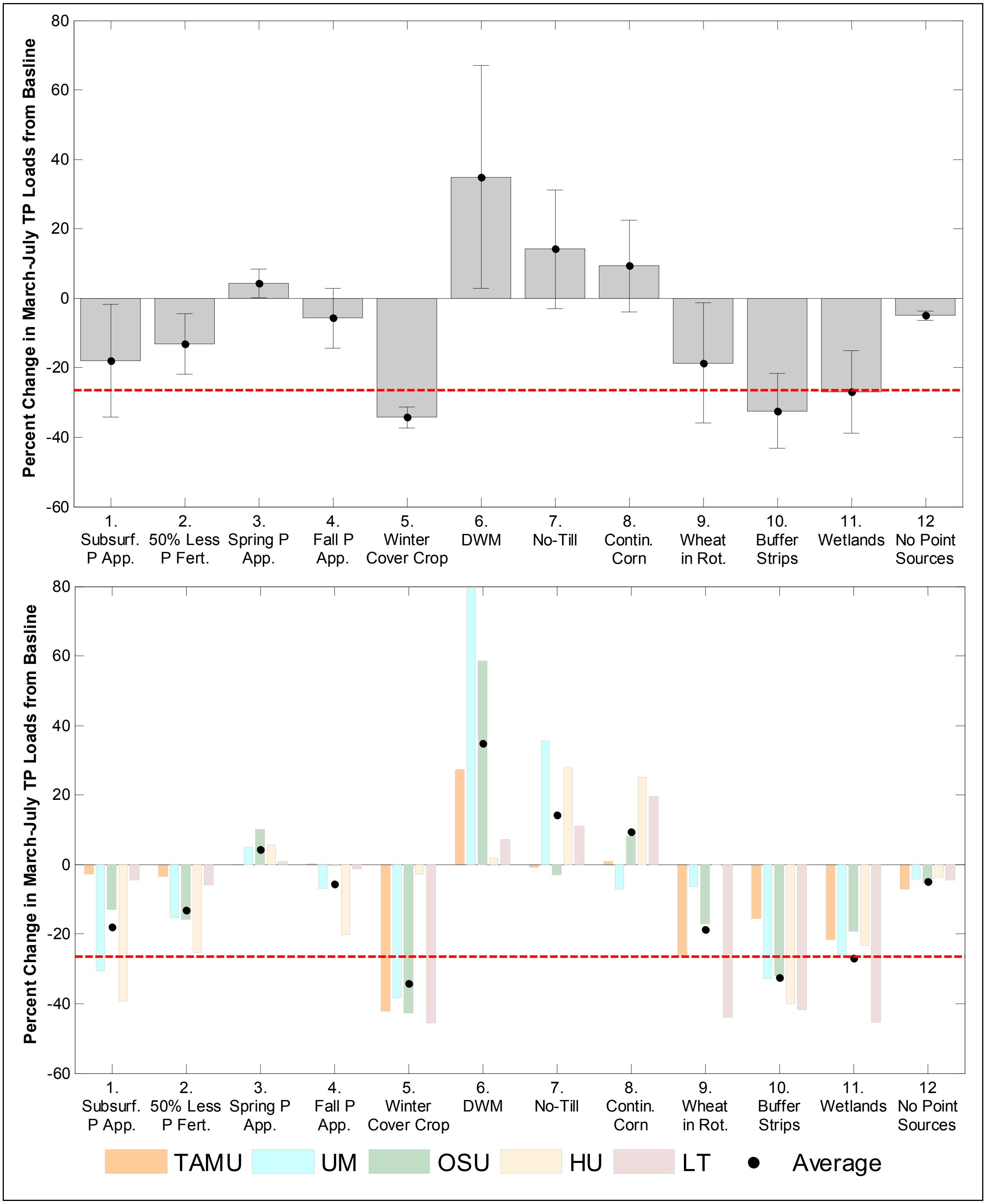


Figure A7.2: Percent change in March-July TP loads from Baseline averaged across all models with standard deviations (top) and for each model (bottom). GLWQA target (red dashed line) estimated as the load target divided by the average load from 2005-2014.

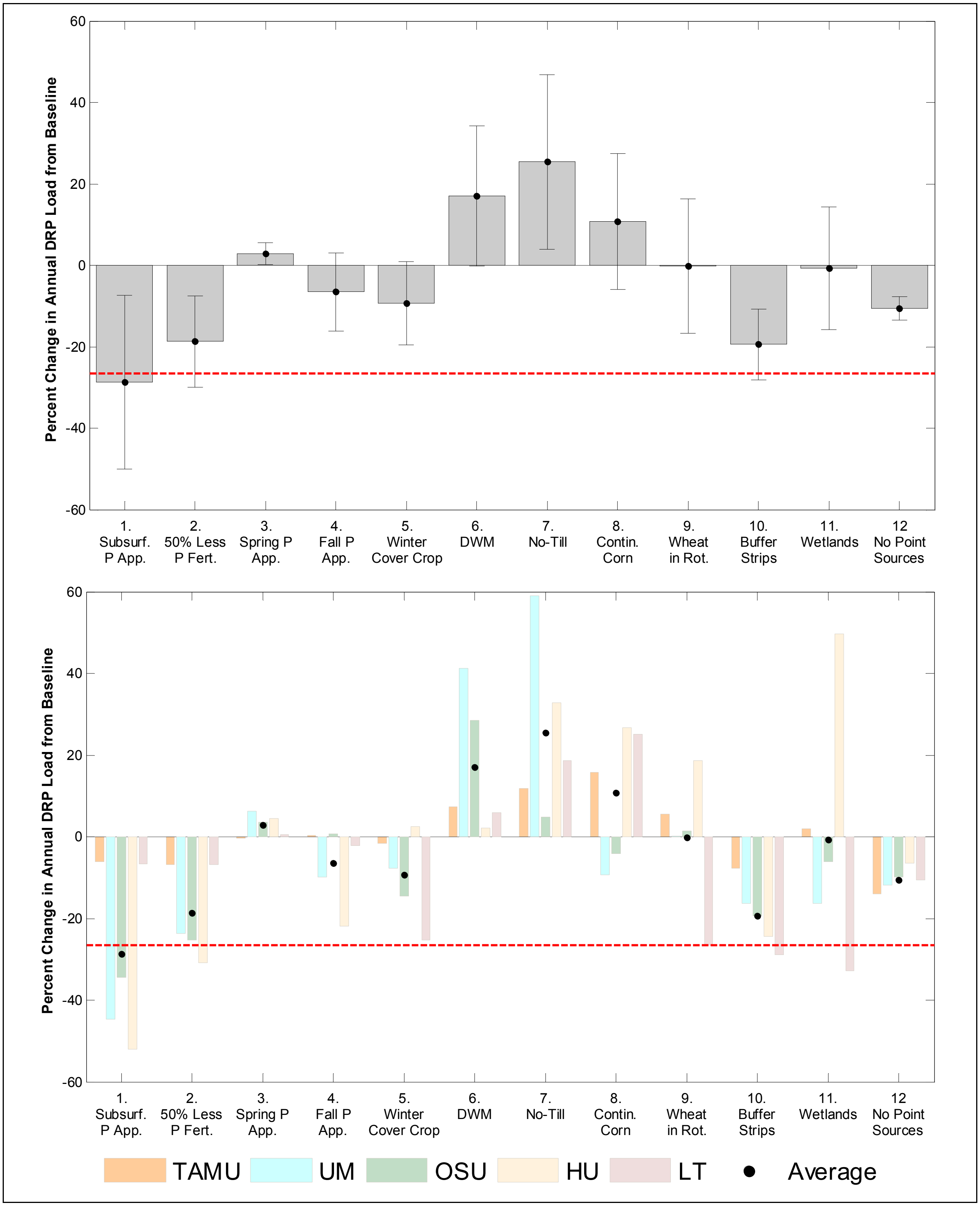


Figure A7.3: Percent change in Annual DRP loads from Baseline averaged across all models with standard deviations (top) and for each model (bottom). GLWQA target (red dashed line) estimated as the load target divided by the average load from 2005-2014.

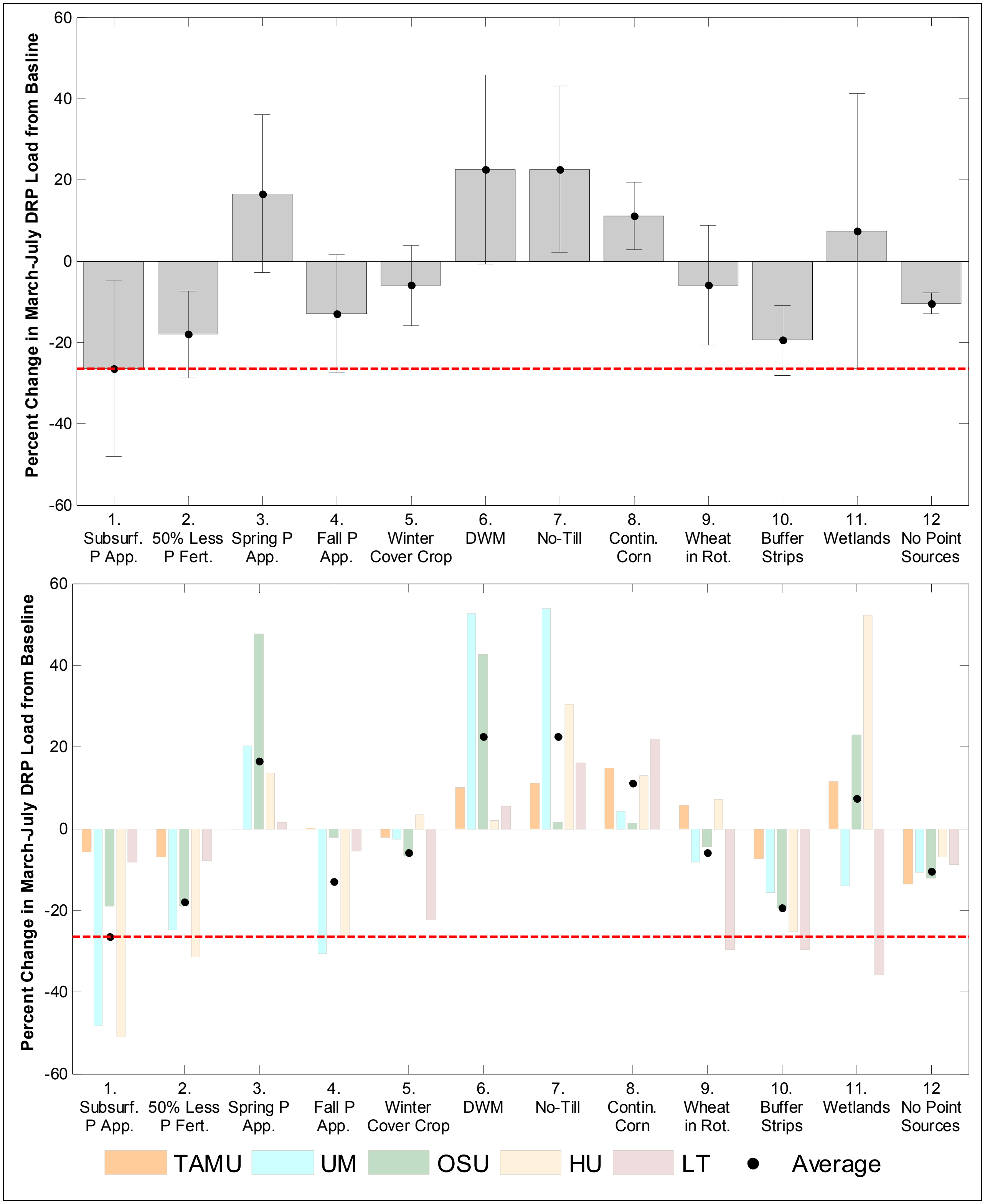


Figure A7.4: Percent change in March-July DRP loads from Baseline averaged across all models with standard deviations (top) and for each model (bottom). GLWQA target (red dashed line) estimated as the load target divided by the average load from 2005-2014.

A8. Bundled scenarios details

Here we provide details on bundled scenarios, including how they were modeled, the extent of cropland taken out of production, impacts on average crop yields, and individual model results for annual TP, DRP, nitrogen, and sediment. Table A8.1 shows the amount of cropland that would be taken out of production or have to change practices in each bundled scenario. Table A8.2 details how each bundle of practices was simulated in the models. Tables A8.3-8.4 show how the simulated scenarios would impact average crop yields in the watershed. Finally, Table A8.5 provides the data for March-July TP and DRP loading as shown in Figure 4.

Table A8.1: Bundled scenario details on the extent of implementation, including the amount of land taken out of production and the amount of land that would have changed management practices. Note that cropland area in the Maumee watershed is estimate at 1,246,800 hectares by extracting "cultivated cropland" in the NLCD (2011) dataset within the Maumee watershed boundaries. †Equivalent to single-practice scenario #12.

#	Bundle name	Adoption area for practices	Cropland taken out of production		Maintained as row cropland but with changed management practices	
		hectares of cropland	*hectares of cropland*	*% of all cropland*	*hectares of cropland*	*% of all cropland*
1†	**No Point Source Discharges**	0	0	0%	0	0%
2a	**Cropland conversion to grassland** at 10% targeted adoption	124,679	124,679	10%	0	0%
2b	**Cropland conversion to grassland** at 25% targeted adoption	311,696	311,696	25%	0	0%
2c	**Cropland conversion to grassland** at 50% targeted adoption	623,393	623,393	50%	0	0%
3	**In-field practices** at 25% random adoption	311,696	0	0%	311,696	25%
4	**Nutrient management** at 25% random adoption	311,696	0	0%	311,696	25%
5	**Nutrient management** at 100% adoption	1,246,785	0	0%	1,246,785	100%
6	**Commonly recommended practices** at 100% random adoption	941,323	6,234	1%	935,089	75%
7	**Continuous no-tillage and subsurface placement of P fertilizer** at 50% random adoption	623,393	0	0%	623,393	50%
8	**Series of practices** at 50% targeted adoption	635,861	12,468	1%	623,393	50%
9	**Series of practices** at 50% random adoption	635,861	12,468	1%	623,393	50%
10	**Diversified rotation** at 50% random adoption	623,393	0	0%	623,393	50%
11	**Wetlands and buffer strips** at 25% targeted adoption	12,884	12,884	1%	0	0%

Reference: National Land Cover Database (NLCD) 2006 (2011 Edition); http://www.mrlc.gov/nlcd06_data.php.

Table A8.2: SWAT modeling details for implementing bundles. †Equivalent to single-practice scenario #12.

#	Bundle name	Modeling details
1†	**No Point Source Discharges**	• All point source discharge effluent and loads were set to zero.
2a-c	**Cropland conversion to grassland** at 10% (2a), 25% (2b), and 50% (2c) targeted adoption	• Targeted first to low crop yielding HRUs (as calculated by SWAT) then by highest TP losses per unit area • Switchgrass was modeled as Shawnee switchgrass (*Panicum vigratum*) based on parameters from Trybula et al. (2014) andwas fertilized once per year with 56 kg N/ha of nitrogen fertilizer; no phosphorus fertilizer was applied. Itwas harvested in October of each year
3	**In-field practices** at 25% random adoption	• 50% reduction in P fertilizer application from Baseline applications • Fall timing of P applications (N remained same as Baseline) • Subsurface placement of P with FRT_SURFACE = 0.01 (1% on soil surface) • Winter cover crop was modeled as cereal rye planted after harvest in winters where no wheat was being grown. Cereal rye was killed before spring planting
4	**Nutrient management** at 25% random adoption	• 50% reduction in P fertilizer application from Baseline applications • Fall P applications (N remained same as Baseline) • Subsurface placement of P with FRT_SURFACE = 0.01 (1% on surface)
5	**Nutrient management** at 100% adoption	• Same as #4
6	**Commonly recommended practices** at 100% random adoption	• 50% reduction in P fertilizer application from Baseline applications • Subsurface placement of P with FRT_SURFACE = 0.01 (1% on soil surface) • Continuous no-tillage with no-till drill at crop planting • Medium-quality buffer strips: Field area to buffer strip ratio = 50, fraction of HRU draining to filter = 0.50, fraction of concentrated flow = 0.25
7	**Continuous no-tillage and subsurface placement of P fertilizer** at 50% random adoption	• Subsurface placement of P with FRT_SURFACE = 0.01 (1% on soil surface) • Continuous no-tillage with no-till drill at crop planting
8	**Series of practices** at 50% targeted adoption	• Targeted to HRUs with highest TP losses per unit area • Subsurface placement of P with FRT_SURFACE = 0.01 (1% on surface) • Medium-quality buffer strips: Field area to buffer strip ratio = 50, fraction of HRU draining to filter = 0.50, fraction of concentrated flow = 0.25 • Winter cover crop was modeled as cereal rye planted after harvest in winters where no wheat was being grown. Cereal rye was killed before spring planting
9	**Series of practices** at 50% random adoption	• Practice details same as in #8, but practice series were randomly applied rather than targeted
10	**Diversified rotation** at 50% random adoption	• A Baseline rotation that included wheat was adopted • Winter cover crop was modeled as cereal rye planted after harvest in winters where no wheat was being grown. Cereal rye was killed before spring planting
11	**Wetlands and buffer strips** at 25% targeted adoption	• Medium performance wetlands were targeted to the 25% of sub-watersheds with the greatest TP loading. Wetlands drained 50% of subbasin area and some models simulated the interception of tile flow while others did not. • Medium-quality buffer strips were targeted to the 25% of HRUs with greatest TP loss: Field area to buffer strip ratio = 50, fraction of HRU draining to filter = 0.50, fraction of concentrated flow = 0.25

Reference: Trybula EM, Cibin R, Burks JL, Chaubey I, Brouder SM, Volenec JJ. 2014. Perennial rhizomatous grasses as bioenergy feedstock in SWAT: parameter development and model improvement. Glob. Change Biol. Bioenerg. 2014, 7(6), 1185-1202; DOI: 10.1111/gcbb.12210.

Table A8.3: The influence of bundled scenarios on ***average crop yields for cropped areas*** in the Maumee watershed. These do not take into account cropland taken out of row crop production. The diversified rotation has little influence on crop yields but is marked *NA* because each model implemented the crop rotation differently, some with double-cropped wheat and soybeans, and results are not easy to interpret. No percent change is given for switchgrass because it was not present in the baseline. Results are the average of the five SWAT models from 2005-2014. †Equivalent to single-practice scenario #12.

#	Bundle name	Watershed average crop yields for cropped areas (t/ha)				Percent change from baseline		
		corn	**soybean**	**wheat**	**switchgrass**	**corn**	**soybean**	**wheat**
0	**Baseline scenario**	8.52	2.46	3.90	0.00	--	--	--
1†	**No Point Source Discharges**	8.52	2.46	3.90	0.00	0%	0%	0%
2a	**Cropland conversion to grassland** at 10% targeted adoption	8.67	2.47	3.90	8.67	2%	0%	0%
2b	**Cropland conversion to grassland** at 25% targeted adoption	8.82	2.47	3.89	8.79	4%	0%	0%
2c	**Cropland conversion to grassland** at 50% targeted adoption	9.09	2.48	3.90	8.91	7%	1%	0%
3	**In-field practices** at 25% random adoption	8.56	2.46	3.90	0.00	1%	0%	0%
4	**Nutrient management** at 25% random adoption	8.51	2.46	3.90	0.00	0%	0%	0%
5	**Nutrient management** at 100% adoption	8.48	2.46	3.87	0.00	0%	0%	-1%
6	**Commonly recommended practices** at 100% random adoption	8.51	2.46	3.89	0.00	0%	0%	0%
7	**Continuous no-tillage and subsurface placement of P fertilizer** at 50% random adoption	8.51	2.46	3.90	0.00	0%	0%	0%
8	**Series of practices** at 50% targeted adoption	8.62	2.45	3.85	0.00	1%	0%	-1%
9	**Series of practices** at 50% random adoption	8.60	2.46	3.91	0.00	1%	0%	0%
10	**Diversified rotation** at 50% random adoption	*NA*	*NA*	*NA*	0.00	*NA*	*NA*	*NA*
11	**Wetlands and buffer strips** at 25% targeted adoption	8.52	2.46	3.90	0.00	0%	0%	0%

Table A8.4: The influence of bundled scenarios on ***total crop yields for the Maumee watershed***, taking into account cropland taken out of row crop production. The diversified rotation has little influence on crop yields but is marked *NA* because each model implemented the crop rotation differently, some with double-cropped wheat and soybeans, and results are not easy to interpret. Note that cropland area in the Maumee watershed is estimate at 1,246,800 hectares by extracting "cultivated cropland" in the NLCD (2011) dataset within the Maumee watershed boundaries. No percent change is given for switchgrass because it was not present in the baseline. Results are the average of the five SWAT models from 2005-2014. †Equivalent to single-practice scenario #12.

#	Bundle name	Total crop yields for the watershed (t)				Percent change from baseline		
		corn	**soybean**	**wheat**	**switchgrass**	**corn**	**soybean**	**wheat**
0	**Baseline scenario**	3,822,897	1,594,176	584,179	0	--	--	--
1†	**No Point Source Discharges**	3,822,897	1,594,176	584,179	0	0%	0%	0%
2a	**Cropland conversion to grassland** at 10% targeted adoption	3,500,978	1,439,110	525,596	1,081,010	-8%	-10%	-10%
2b	**Cropland conversion to grassland** at 25% targeted adoption	2,970,569	1,199,566	436,864	2,738,766	-22%	-25%	-25%
2c	**Cropland conversion to grassland** at 50% targeted adoption	1,984,108	796,573	291,240	5,551,490	-48%	-50%	-50%
3	**In-field practices** at 25% random adoption	3,826,504	1,594,108	584,256	0	0%	0%	0%
4	**Nutrient management** at 25% random adoption	3,819,707	1,594,185	583,452	0	0%	0%	0%
5	**Nutrient management** at 100% adoption	3,806,474	1,594,074	579,303	0	0%	0%	-1%
6	**Commonly recommended practices** at 100% random adoption	3,799,436	1,586,288	578,981	0	-1%	0%	-1%
7	**Continuous no-tillage and subsurface placement of P fertilizer** at 50% random adoption	3,820,745	1,594,389	583,227	0	0%	0%	0%
8	**Series of practices** at 50% targeted adoption	3,829,237	1,575,361	570,861	0	0%	-1%	-2%
9	**Series of practices** at 50% random adoption	3,820,000	1,577,393	578,870	0	0%	-1%	-1%
10	**Diversified rotation** at 50% random adoption	*NA*	*NA*	*NA*	0	*NA*	*NA*	*NA*
11	**Wetlands and buffer strips** at 25% targeted adoption	3,783,392	1,577,703	578,142	0	-1%	-1%	-1%

Reference: National Land Cover Database (NLCD) 2006 (2011 Edition); http://www.mrlc.gov/nlcd06_data.php.

Table A8.5: March-July TP and DRP loading for each scenario, as shown in Figure 4. [†]Equivalent to single-practice scenario #12.

#	Bundle name	TP loading (tonnes)		DRP loading (tonnes)	
		2005-2014 average		2005-2014 average	
0	**Baseline scenario**	1,120		242	
		5-model mean	*standard deviation*	**5-model mean**	*standard deviation*
1[†]	**No Point Source Discharges**	1,065	*16*	217	*6*
2a	**Cropland conversion to grassland** at 10% targeted adoption	954	*66*	222	*13*
2b	**Cropland conversion to grassland** at 25% targeted adoption	782	*99*	200	*24*
2c	**Cropland conversion to grassland** at 50% targeted adoption	567	*113*	171	*41*
3	**In-field practices** at 25% random adoption	985	*36*	229	*15*
4	**Nutrient management** at 25% random adoption	1,054	*50*	223	*13*
5	**Nutrient management** at 100% adoption	870	*198*	167	*55*
6	**Commonly recommended practices** at 100% random adoption	880	*173*	189	*46*
7	**Continuous no-tillage and subsurface placement of P fertilizer** at 50% random adoption	1021	*77*	218	*28*
8	**Series of practices** at 50% targeted adoption	639	*75*	176	*28*
9	**Series of practices** at 50% random adoption	764	*49*	189	*21*
10	**Diversified rotation** at 50% random adoption	888	*84*	228	*21*
11	**Wetlands and buffer strips** at 25% targeted adoption	771	*189*	209	*28*

This section also provides March-July TP and DRP load comparisons across models (Figure A8.1) as well as percent change comparisons for other outputs intended to provide information on how the bundles may impact other water quality parameters (Figures A8.2-A8.4). Note that not all modelers prioritized calibrating nitrogen and sediment (see calibration variables by model in Table A2.1), so the results for these two parameters should be interpreted within this context.

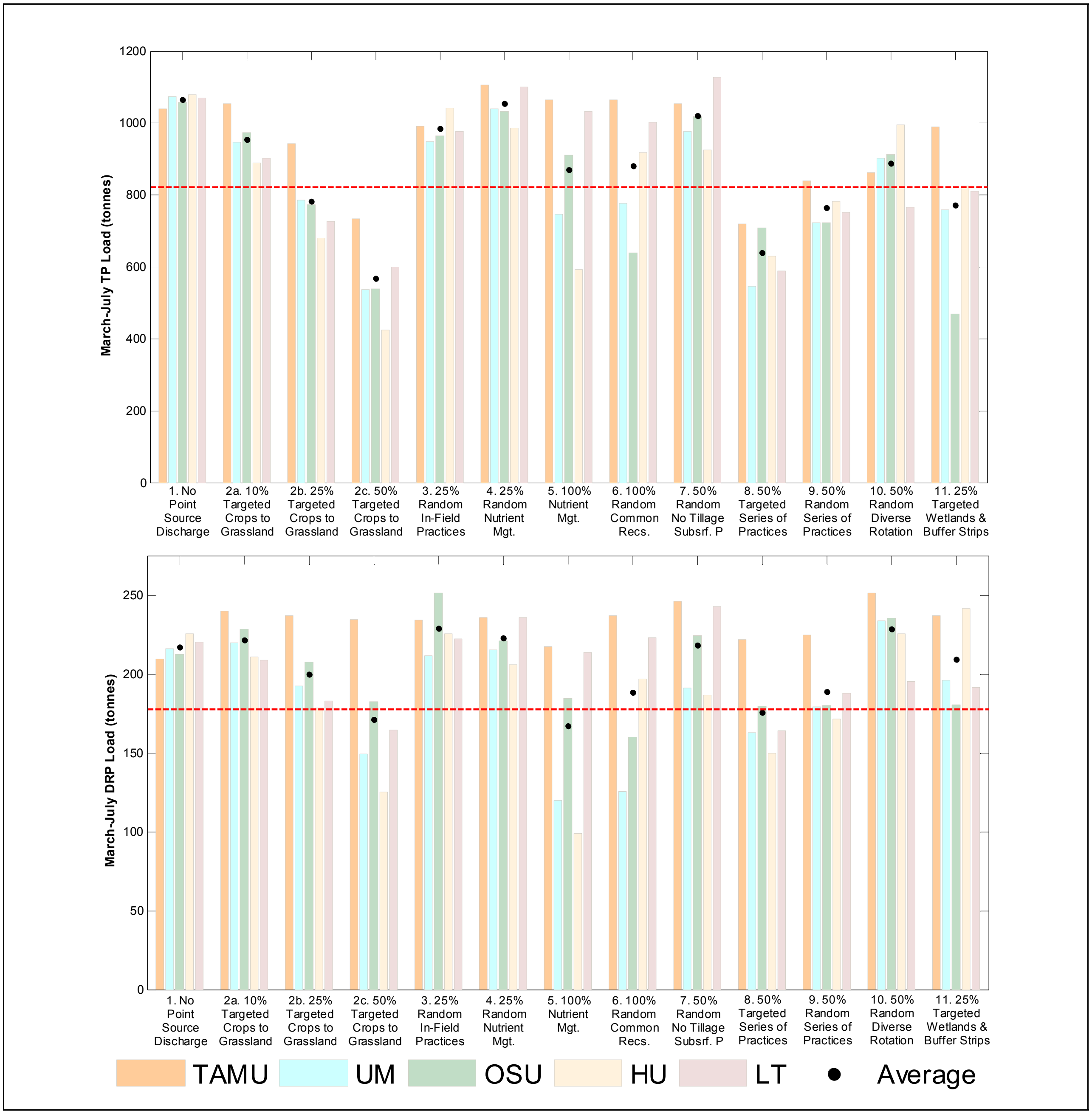


Figure A8.1: March-July TP (top) and DRP (bottom) loads from Baseline across all models. GLWQA target loads are shown by the red dashed lines. Model biases were removed from these loads by calculating the percent change between each model's Baseline and scenario and then applying that percent change to the average observed loads during the time period.

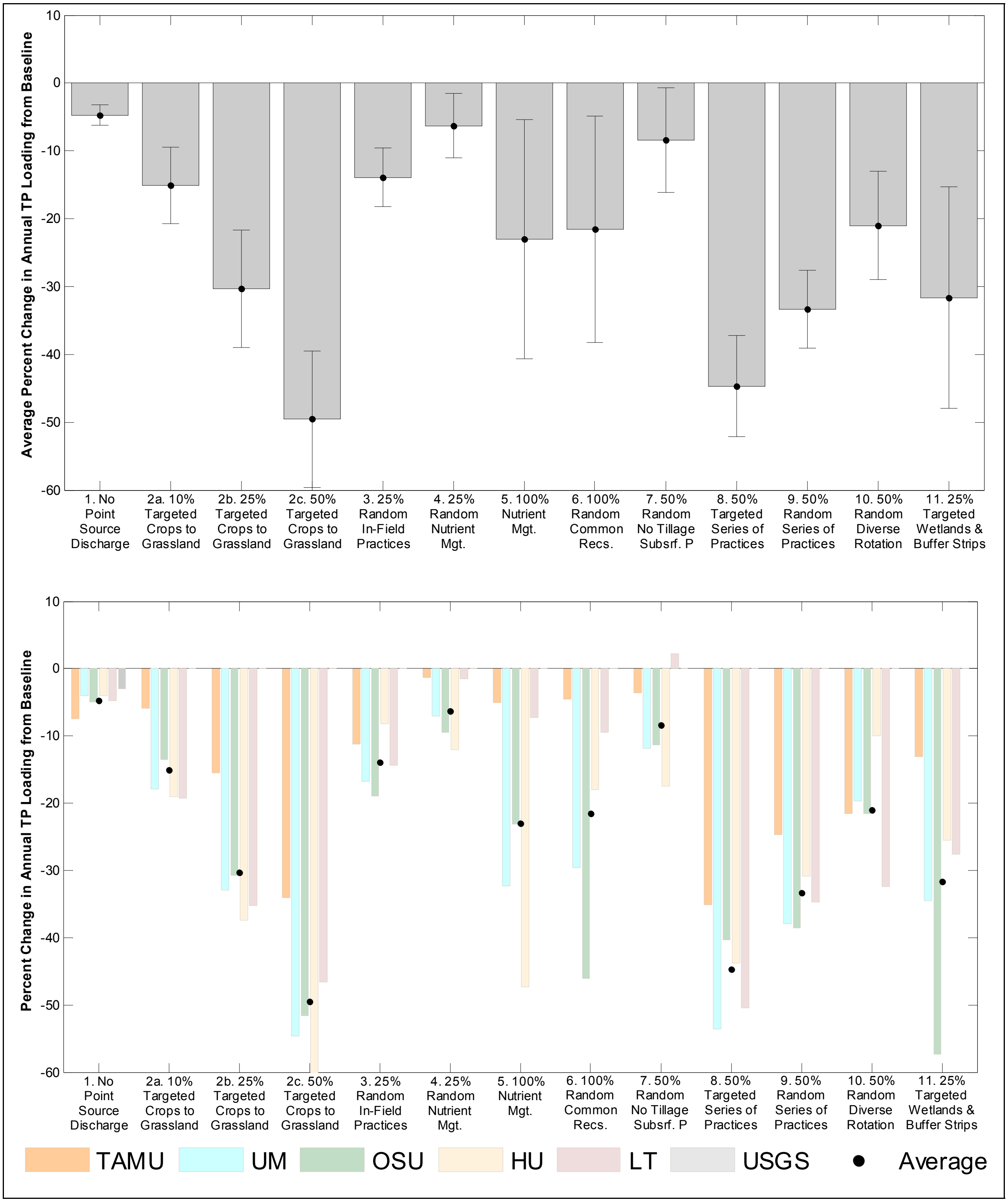


Figure A8.2: Percent change in annual TP loads from Baseline averaged across all models with standard deviations (top) and for each model with average (bottom). The SPARROW (USGS) model was only able to run scenario #1 (no point source discharge) and therefore scenarios #2-11 only include the 5 SWAT model results.

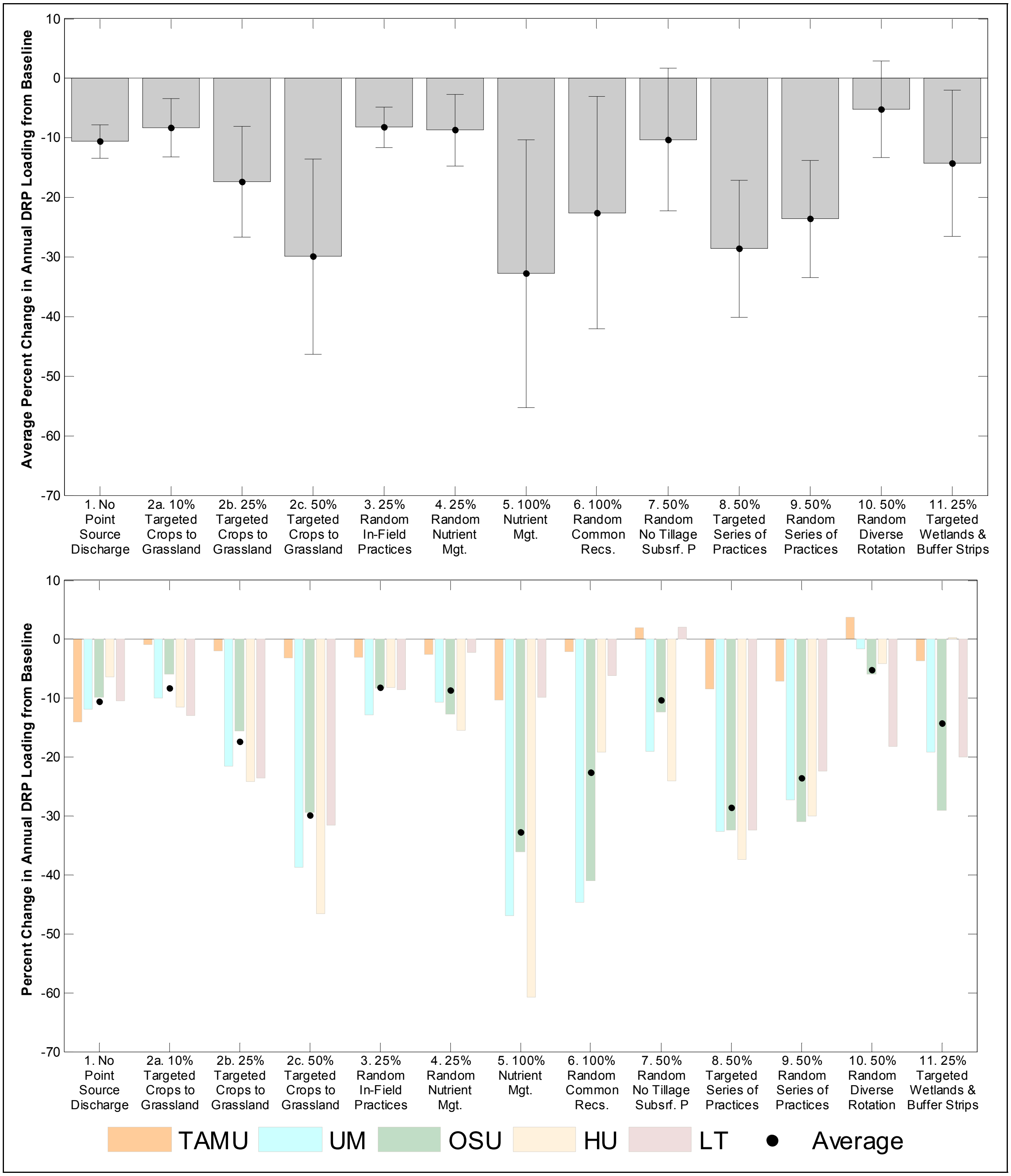


Figure A8.3: Percent change in annual dissolved reactive phosphorus (DRP) loads from Baseline averaged across all models with standard deviations (top) and for each model with average (bottom).

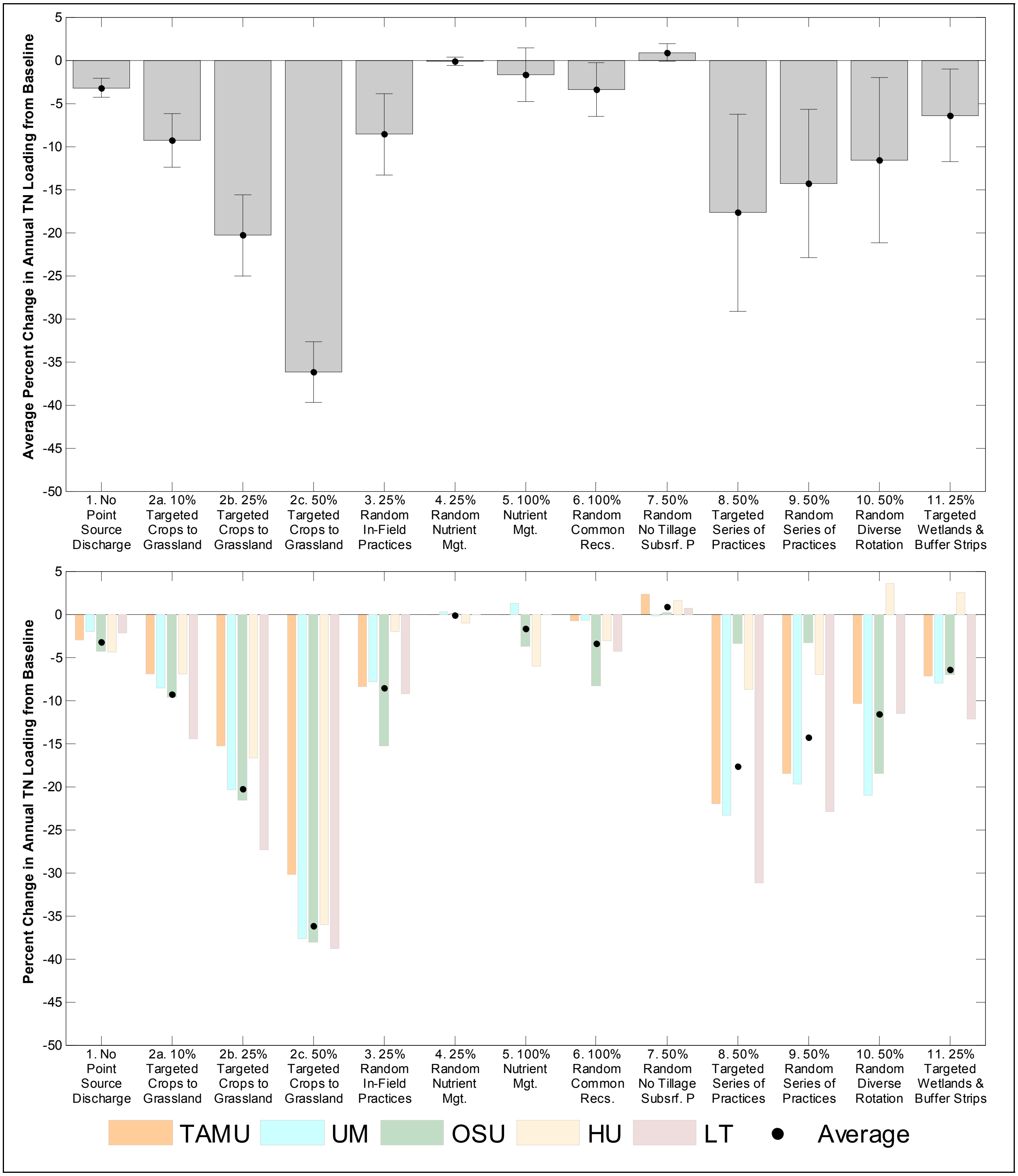


Figure A8.4: Percent change in annual total nitrogen (TN) loads from Baseline averaged across all models with standard deviations (top) and for each model with average (bottom).

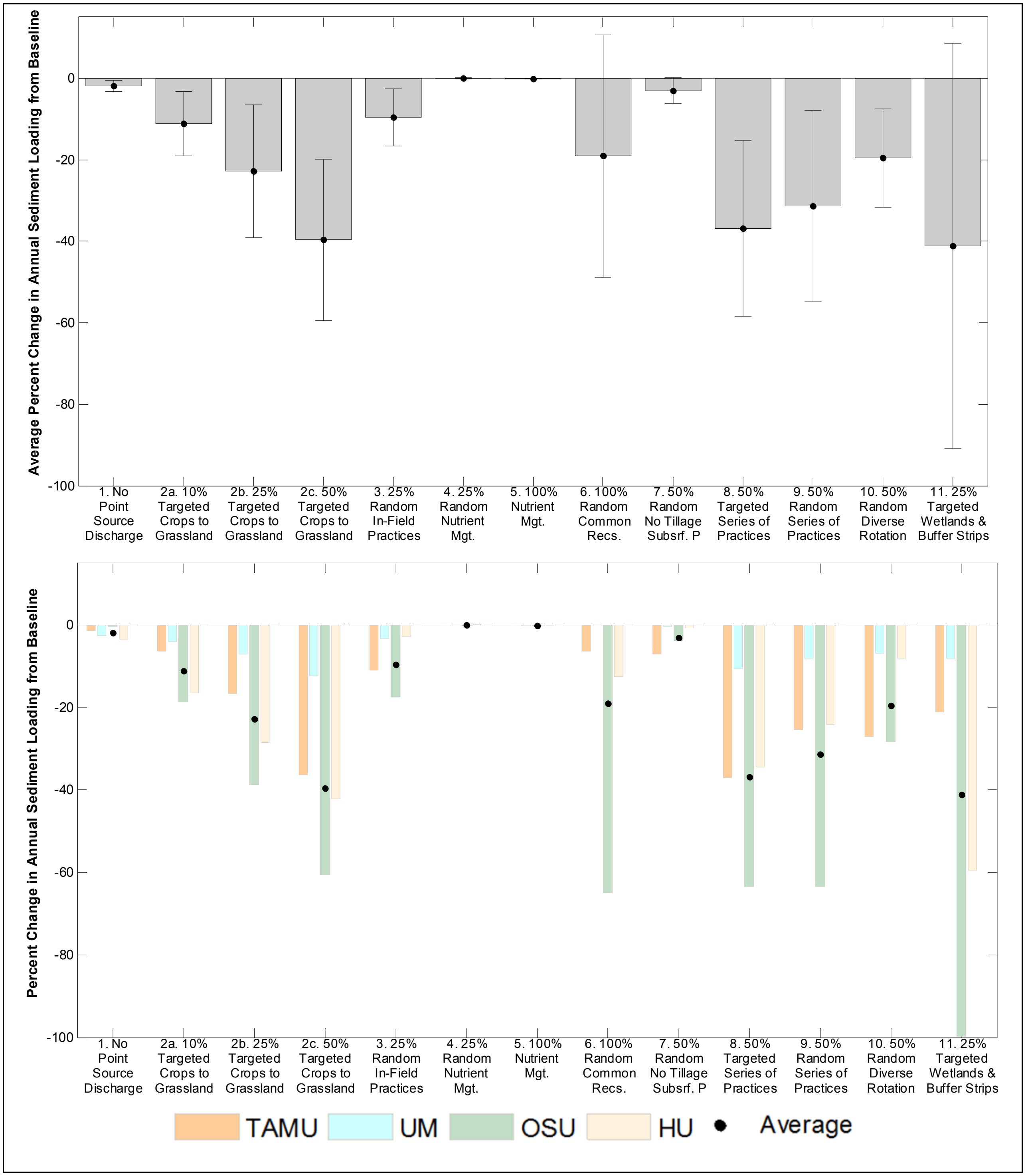


Figure A8.5: Percent change in annual sediment loads from Baseline averaged across all models with standard deviations (top) and for each model with average (bottom). The LimnoTech model was not calibrated for sediment at the start of the project, so the results are not included here.

A9. Potential Hotspot Identification Methods

The five SWAT models and SPARROW were used to describe the areas in the Maumee River watershed having the highest TP and DRP yields (kg/ha/y) delivered to the lake (Figures A9.1 & A9.2).

Potential Hotspot Identification Method in SWAT

To identify hotspots (vulnerable areas) with the SWAT model results, we started by extracting information from the river shapefile (riv.shp), the main channel output file (output.rch), and the reservoir output file (output.rsv) to determine a flow and nutrient routing sequence. In the river shapefile, the *Subbasin* and *SubbasinR* fields from the attribute table show the upstream and downstream reach number, respectively. Flow and nutrient of a sub-watershed in *Subbasin* was transported to the one in *SubbasinR*. Based on this, we constructed a network of flow and nutrient transport sequence for all sub-watersheds.

Based on the output.rch file, total phosphorus transported into each sub-watershed (*total phosphorus in*) was calculated by summing the organic phosphorus transported into (data in *ORGP_IN* field) and mineral phosphorus transported into (data in *MINP_IN* field) that sub-watershed. Similarly, total phosphorus transported out of each sub-watershed (total phosphorus out) was calculated by summing the organic phosphorus transported out of (data in *ORGP_OUT* field) and mineral phosphorus transported out of (data in *ORGP_OUT* field) that sub-watershed. After that, we then subtracted *total phosphorus out* of the upstream sub-watershed from *total phosphorus in* of downstream sub-watershed, following the network of flow and nutrient transport sequence we built. If there were reservoir loads in a sub-watershed, the loads, calculated from output.rsv, was also subtracted *from total phosphorus out*. The monthly average of the loads was further summarized to annual average. The results represent the amount of total phosphorus originating from each sub-watershed.

The sources of phosphorus loads to each sub-watershed may include upstream sub-watershed(s), reservoirs, point sources, and the contribution from the land in each sub-watershed itself. To calculate the percentage of loads from these sources to each sub-watershed, these loads were divided by the total load from that watershed. After this step, we calculated the loads delivered to the lake from each source by multiplying the percentage by the total load delivered based on the source percentage calculation and the network of flow and nutrient transport sequence. The calculation started in the most downstream sub-watershed and then moved toward the upstream sub-watershed(s) until all the sub-watersheds were calculated. Maps were then created using quantile classification of delivered yield (kg/ha/y) (Figure A9.1). We calculated the delivered loads of DRP from each sub-watershed using the same above-mentioned approach for TP loads, except that the calculation only includes data from *MINP_IN* and *MINP_OUT* (Figure A9.2).

Potential Hotspot Identification Method in SPARROW

All of the catchments in the SPARROW model were ranked based on their relative delivered incremental yields to the outlet of the Maumee River Basin (Figure A9.1).

Notes on Differences in Hotspots

Differences among the hotspot maps can be attributed to varying assumptions about the spatial locations of practices such as tile drainage, subsurface or incorporated phosphorus applications, tillage practices, and fertilizer rates. Specifically, some differences between the SWAT models and the SPARROW model arise from SPARROW's elevated use of manures in the southern portion of the watershed and the fact that SPARROW assumes a higher delivery ratio for manure than inorganic fertilizers, whereas SWAT appears to treat their transport more equally.

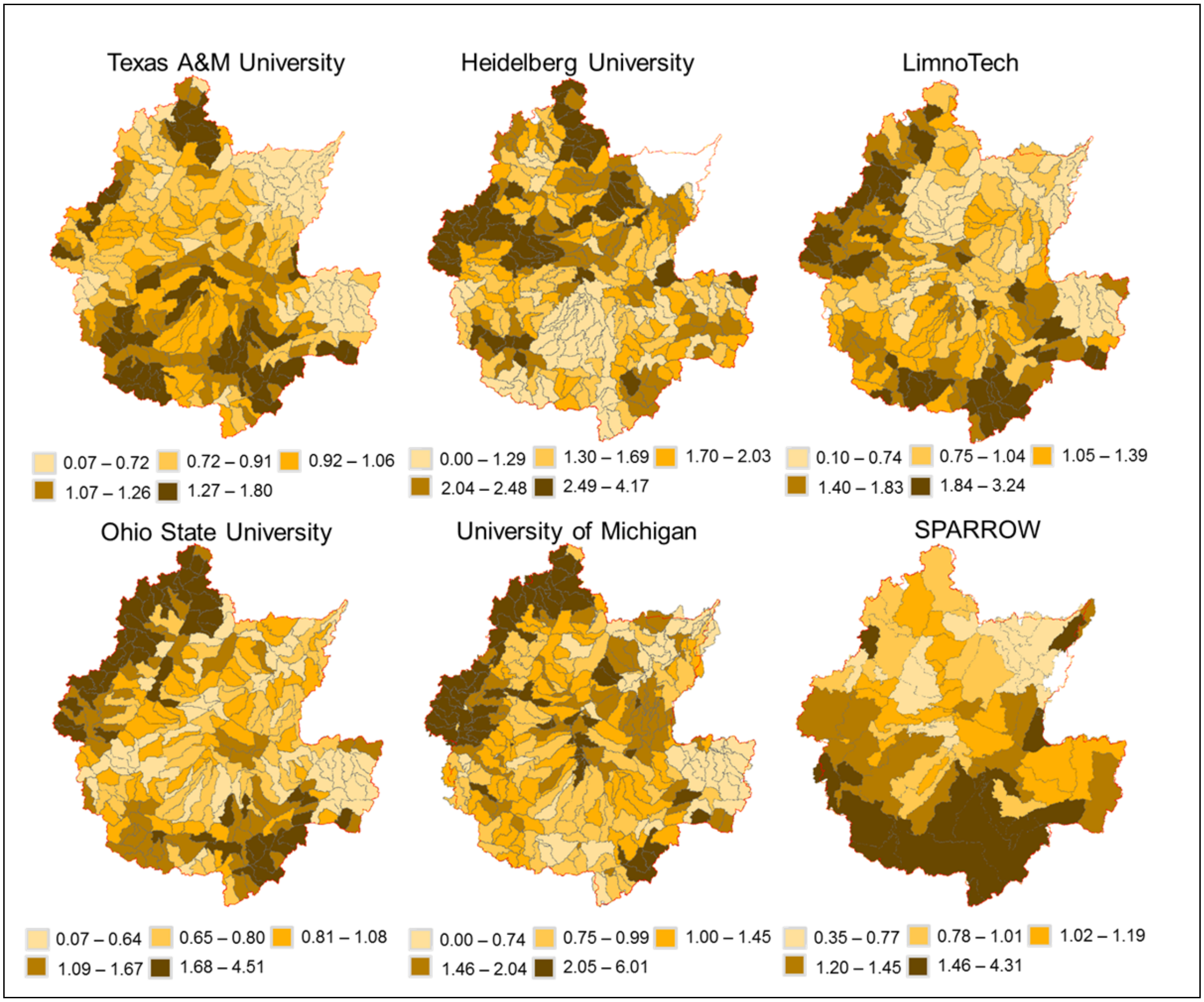


Figure A9.1: TP sub-watershed physically vulnerable areas or "hotspots", in terms of yields (kg/ha/y). The red outline highlights the USGS HUC extent for the MRW, emphasizing the spatial differences between models based on setup. Hotspots are represented by the darkest colors shown in each legend.

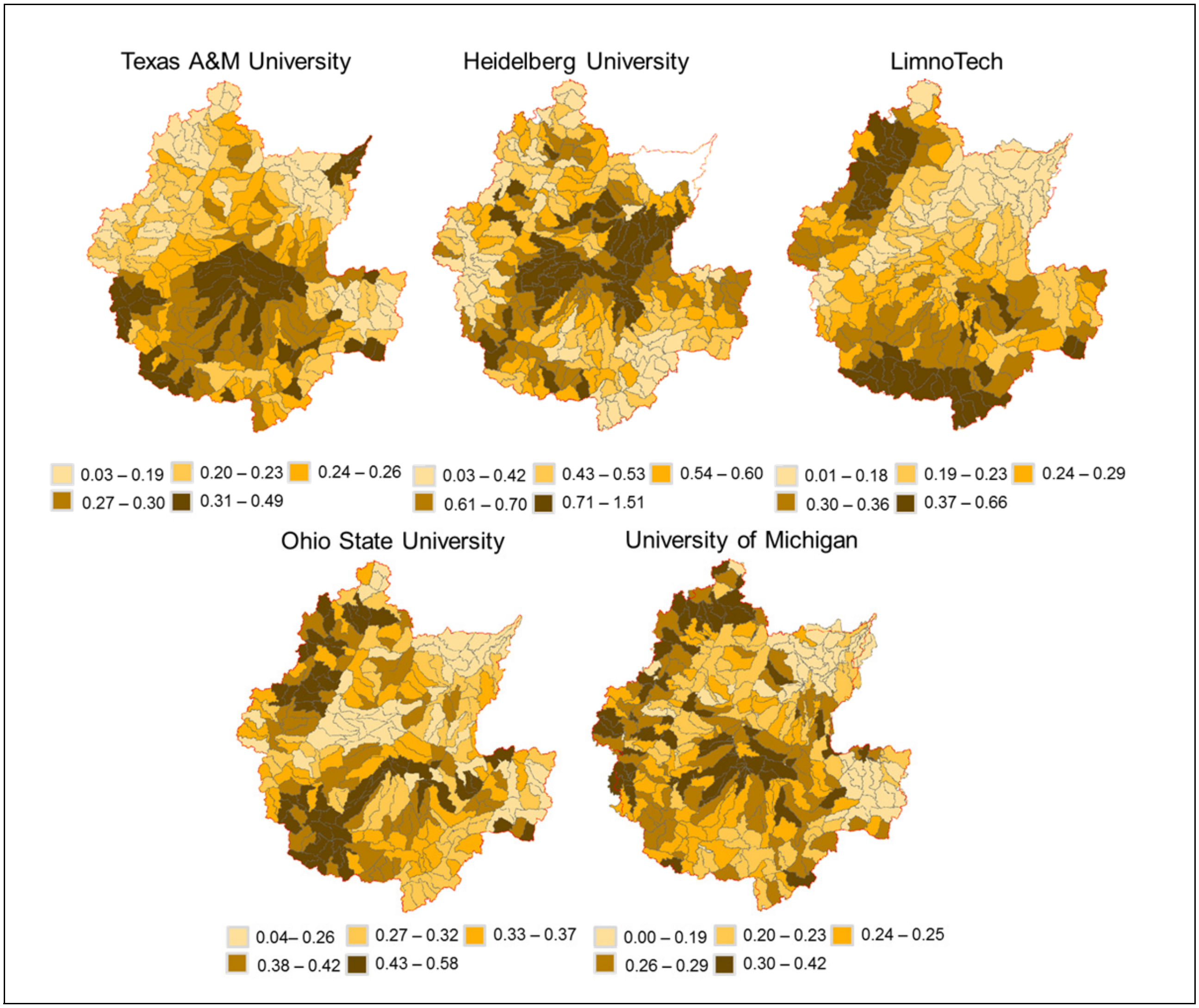


Figure A9.2: Dissolved reactive phosphorus sub-watershed physically vulnerable areas or "hotspots" (kg/ha/y). The red outline highlights the USGS HUC extent for the MRW, emphasizing the spatial differences between models based on setup. The SPARROW model only outputs TP, therefore is not included in this figure.

A10. Baseline validation results

All SWAT models were validated to monthly data at Waterville, OH during the 2005-2014 time period (Table A10.1). Additional calibration checks were performed to compare the model performance to other historical data (Table A10.2). Annual time series plots for average streamflow and loads of TP, TN, and sediment are also provided (Figures A10.1-A10.5). *NA* indicates a model was not calibrated to a constituent and so results are not reported.

Table A10.1: Validation to streamflow and water quality at the USGS and NCWQR stations near Waterville, OH.

		Heidelberg	LimnoTech	OSU	TAMU	UM	Average
	NS	0.82	0.90	0.91	0.86	0.89	0.88
Flow (cms)	PBIAS	-7.18	9.53	9.97	11.40	5.58	5.86
	R^2	0.86	0.91	0.93	0.88	0.91	0.90
	NS	0.64	0.82	0.73	0.56	0.70	0.69
TP (kg)	PBIAS	36.72	-5.56	-6.63	-22.17	6.94	1.86
	R^2	0.74	0.82	0.75	0.71	0.70	0.75
	NS	-0.02	0.71	0.51	0.52	0.46	0.44
DRP (kg)	PBIAS	80.76	1.48	16.11	-13.33	-12.76	14.45
	R^2	0.55	0.71	0.54	0.70	0.51	0.60
	NS	0.55	*NA*	0.69	0.70	0.87	0.70
Sediment (tonnes)	PBIAS	-28.59	*NA*	-4.59	30.05	11.15	-2.01
	R^2	0.69	*NA*	0.69	0.76	0.88	0.76
	NS	0.39	0.54	0.23	0.22	0.73	0.42
TN (kg)	PBIAS	-29.51	15.97	-52.45	19.58	3.74	-8.54
	R^2	0.55	0.75	0.58	0.62	0.77	0.66
	NS	0.10	0.21	0.42	-0.59	0.39	0.10
Nitrate (kg)	PBIAS	-7.37	21.98	-37.88	31.30	5.79	2.76
	R^2	0.51	0.65	0.57	0.55	0.62	0.58

Table A10.2: Additional calibration checks (i.e., ensuring that outputs that do not have associated observed data are within known ranges) for each model. Values are reported for the entire watershed, therefore the nutrient section includes all lands (not just agricultural lands) in the per area calculation. However, the crop yields are only reported per agricultural land area. Example references provide the range of outputs that might be expected from the literature; *NA* indicates a range is not applicable or unknown.

Hydrology (mm)						
	Ex. Refs.	HU	LT	OSU	TAMU	UM
Precipitation	NOAA	970	976	973	966	975
Snow fall	NOAA	75	106	106	103	61
Surface runoff	*NA*	275	195	191	220	166
Tile flow	*NA*	50	139	110	91	135
Evapotranspiration	*NA*	632	571	567	593	598
Potential Evapotranspiration	*NA*	1052	1009	1045	1047	1092
Nutrients (kg/ha)						
	Ex. Refs.	HU	LT	OSU	TAMU	UM
Soluble P through tiles	King et al. 2015	0.283	0.068	0.157	0.046	0.083
N fertilizer applied	USDA-ERS.;	66	59	63	55	82
P fertilizer applied	Ruddy et al.	22	12	21	10	13
Organic P in fertilizer	2006	0	0.16	0	0	1.46
Initial mineral P in soil	*NA*	4,559	3,708	7,810	14,660	894
Final mineral P in soil	*NA*	4,566	3,623	7,764	14,546	793
Initial Organic P in soil	*NA*	321	1676	33	1846	1455
Final Organic P in soil	*NA*	306	1705	89	1890	1526
Δ mineral P in soil	*NA*	7	-85	-46	-114	-101
Δ organic P in soil	*NA*	-15	29	56	44	71
Crop Yields (t/ha)						
	Ex. Refs.	HU	LT	OSU	TAMU	UM
Corn	NASS Survey	8.7	7.5	8.3	9.1	9.5
Soybean	& Census	2.4	2.6	2.5	2.2	2.5
Wheat	data	4.1	2.4	4.2	3.8	5.0

References:

- King KW, Williams MR, Macrae ML, Fausey NR, Frankenberger J, Smith DR, Kleinman PJ, Brown LC. 2015. Phosphorus transport in agricultural subsurface drainage: A review. Journal of environmental quality, 44(2):467-85.
- Ruddy BC, Lorenz DL, Mueller, DK. 2006. County-level estimates of nutrient inputs to the land surface of the conterminous United States, 1982-2001. USGS Scientific Investigations Report 2006-5012. Available online at: http://pubs.usgs.gov/sir/2006/5012/.

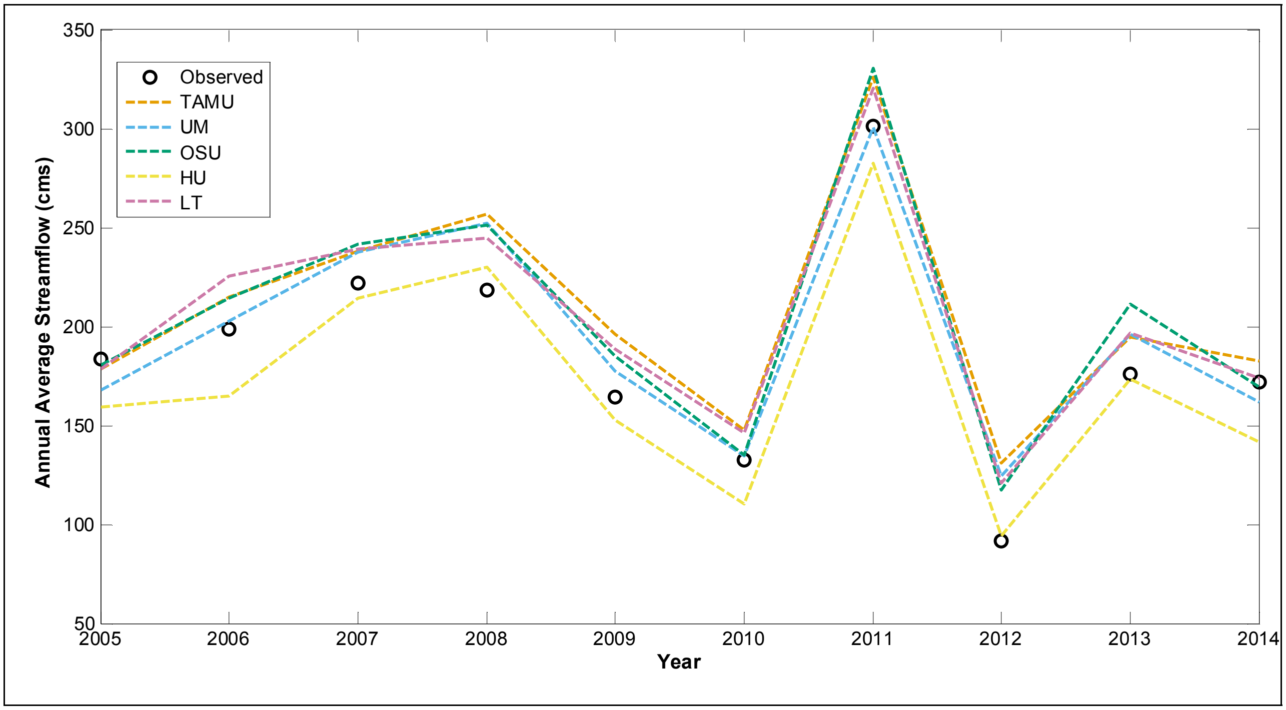


Figure A10.1: Annual average streamflow simulated across each model and compared with filled-in observed data.

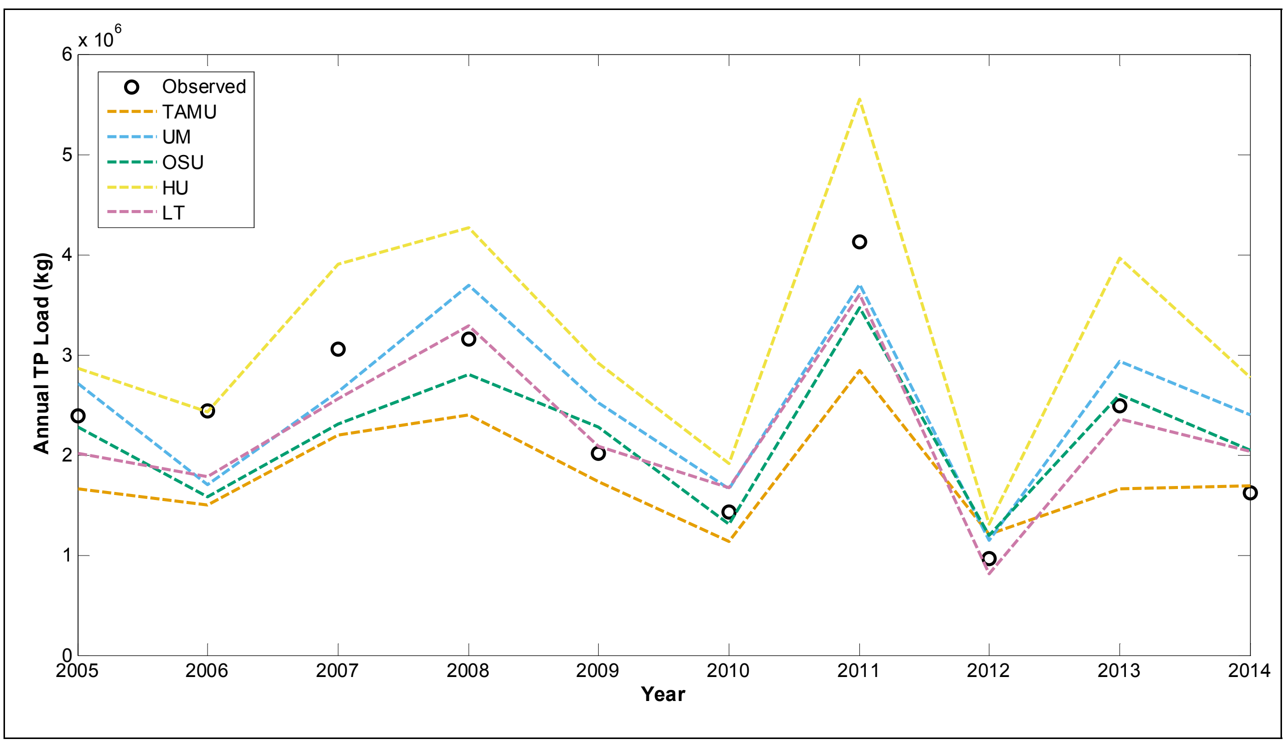


Figure A10.2: Annual TP loads simulated across each model and compared with filled-in observed data.

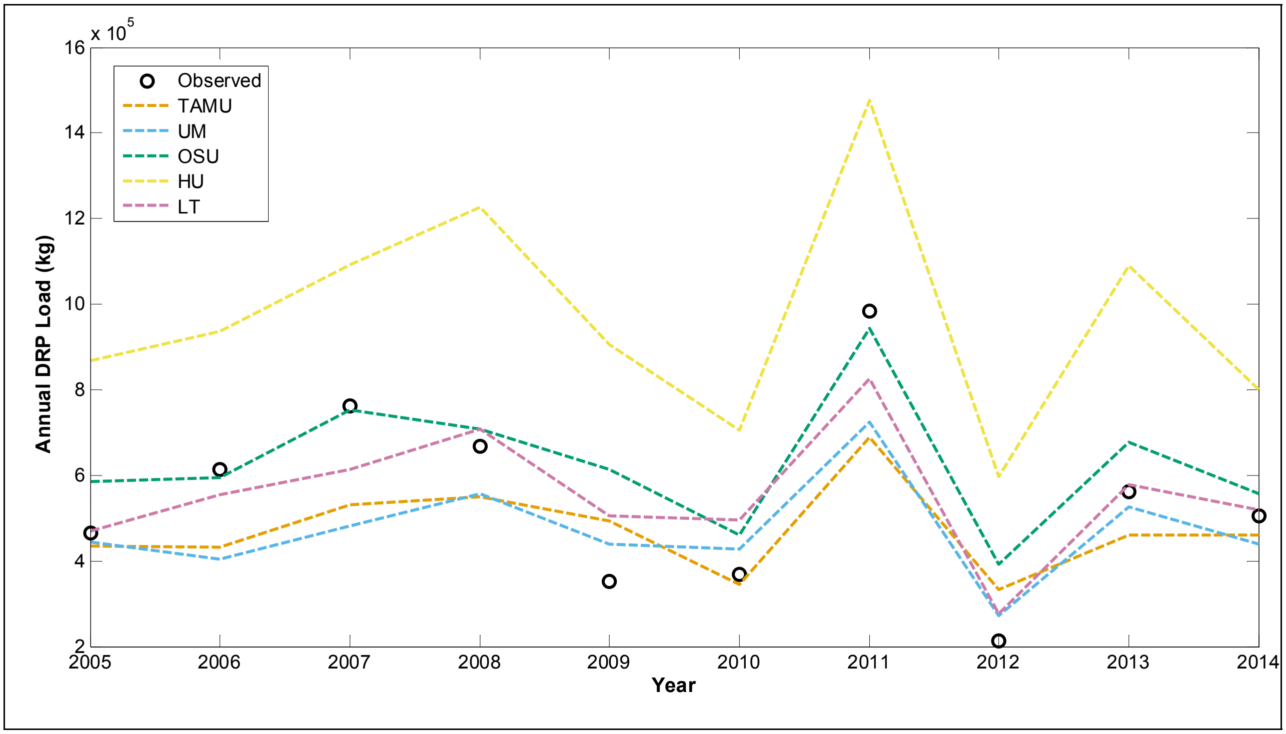


Figure A10.3: Annual DRP loads simulated across each model and compared with filled-in observed data.

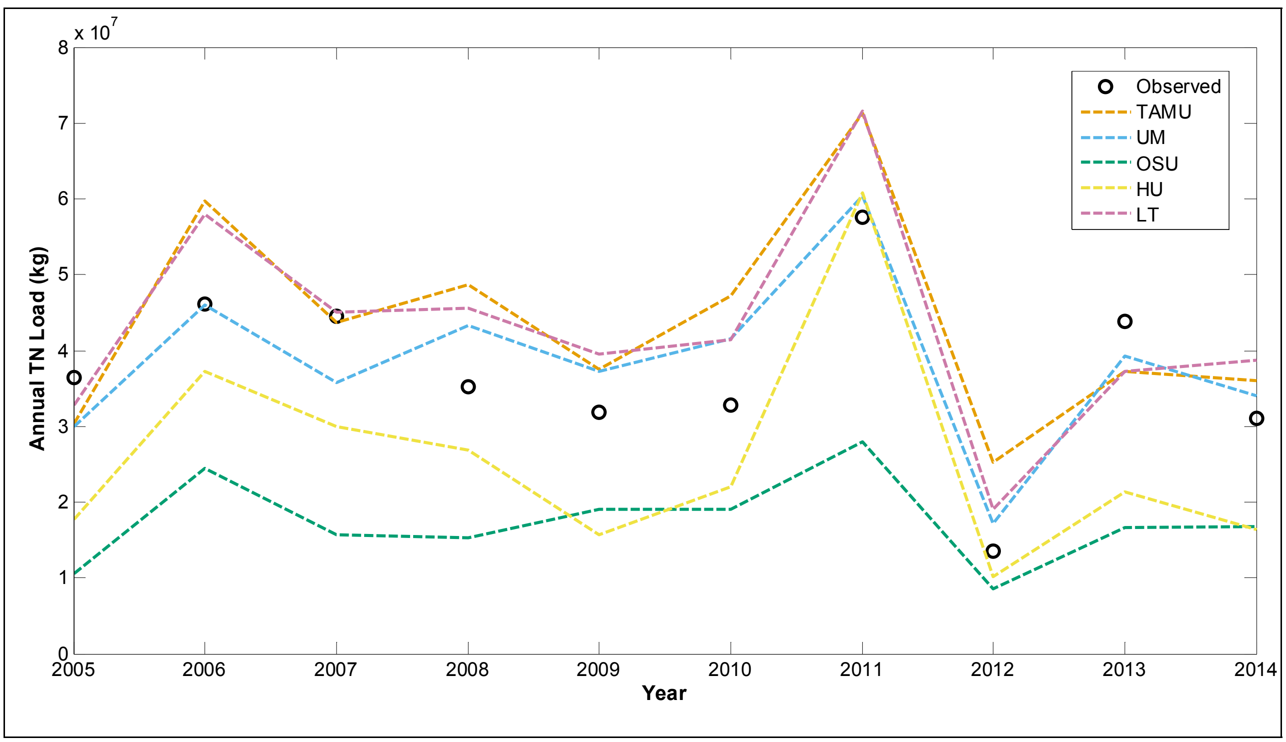


Figure A10.4: Annual TN loads simulated across each model and compared with filled-in observed data.

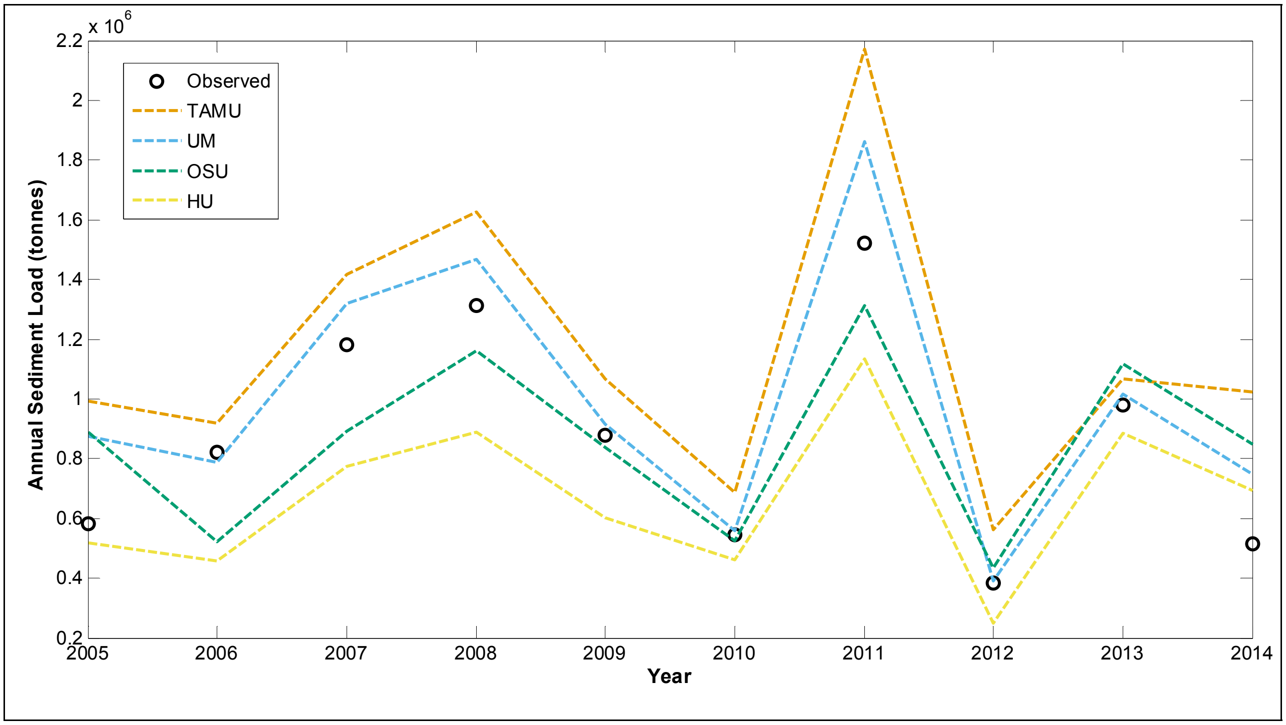


Figure A10.5: Annual sediment loads simulated across each model and compared with filled-in observed data. The LimnoTech model was not calibrated for sediment at the start of the project, so the results are not included here.

GRAHAM
SUSTAINABILITY INSTITUTE
UNIVERSITY OF MICHIGAN

The Water Center engages researchers, practitioners, policymakers, and nonprofit groups to support, integrate, and improve current and future freshwater restoration and protection efforts. The Water Center conducts collaborative science, supporting Great Lakes restoration and coordinates the National Estuarine Research Reserve System (NERRS) Science Collaborative. The Water Center is part of the U-M Graham Sustainability Institute, which fosters sustainability through translational knowledge, transformative learning, and institutional leadership.