ARTICLE IN PRESS

Journal of Environmental Management xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Environmental Management

journal homepage: http://www.elsevier.com/locate/jenvman

Research article

Uncertainty in critical source area predictions from watershed-scale hydrologic models

Grey R. Evenson [a,*,1], Margaret Kalcic [a,b], Yu-Chen Wang [c], Dale Robertson [d], Donald Scavia [c], Jay Martin [a,e], Noel Aloysius [f], Anna Apostel [a], Chelsie Boles [g], Michael Brooker [a], Remegio Confesor [h], Awoke Teshager Dagnew [i], Tian Guo [h], Jeffrey Kast [a], Haley Kujawa [a], Rebecca Logsdon Muenich [j], Asmita Murumkar [a], Todd Redder [g]

[a] Department of Food, Agricultural and Biological Engineering, The Ohio State University, Columbus, OH, USA
[b] The Ohio State University Translational Data Analytics Institute, Columbus, OH, USA
[c] School for Environment and Sustainability, University of Michigan, Ann Arbor, MI, USA
[d] U.S. Geological Survey, Upper Midwest Water Science Center, Middleton, WI, USA
[e] The Ohio State University Sustainability Institute, Columbus, OH, USA
[f] Department of Biomedical, Biological and Chemical Engineering, and School of Natural Resources, University of Missouri, Columbia, MO, USA
[g] LimnoTech, Ann Arbor, MI, USA
[h] Heidelberg University, Tiffin, OH, USA
[i] Environmental Consulting and Technology, Inc., Ann Arbor, MI, USA
[j] School of Sustainable Engineering and the Built Environment, Arizona State University, Tempe, AZ, USA

ARTICLE INFO

Keywords:
Hotspots
Targeting
Prioritization
Best management practices, (BMPs)
SWAT
SPARROW

ABSTRACT

Watershed-scale hydrologic models are frequently used to inform conservation and restoration efforts by identifying critical source areas (CSAs; alternatively 'hotspots'), defined as areas that export relatively greater quantities of nutrients and sediment. The CSAs can then be prioritized or 'targeted' for conservation and restoration to ensure efficient use of limited resources. However, CSA simulations from watershed-scale hydrologic models may be uncertain and it is critical that the extent and implications of this uncertainty be conveyed to stakeholders and decision makers. We used an ensemble of four independently developed Soil and Water Assessment Tool (SWAT) models and a SPAtially Referenced Regression On Watershed attributes (SPARROW) model to simulate CSA locations for flow, phosphorus, nitrogen, and sediment within the ~17,000-km^2 Maumee River watershed at the HUC-12 scale. We then assessed uncertainty in CSA simulations determined as the variation in CSA locations across the models. Our application of an ensemble of models - differing with respect to inputs, structure, and parameterization - facilitated an improved accounting of CSA prediction uncertainty. We found that the models agreed on the location of a subset of CSAs, and that these locations may be targeted with relative confidence. However, models more often disagreed on CSA locations. On average, only 16%–46% of HUC-12 subwatersheds simulated as a CSA by one model were also simulated as a CSA by a different model. Our work shows that simulated CSA locations are highly uncertain and may vary substantially across models. Hence, while models may be useful in informing conservation and restoration planning, their application to identify CSA locations would benefit from comprehensive uncertainty analyses to avoid inefficient use of limited resources.

Credit author statement

Grey R. Evenson: Conceptualization, Methodology, Formal analysis, Investigation, Writing - original draft, Writing - review & editing, Visualization. Margaret Kalcic: Conceptualization, Methodology, Formal analysis, Investigation, Writing - review & editing, Visualization,

* Corresponding author.
E-mail address: evenson.grey@epa.gov (G.R. Evenson).

[1] Present Address: Center for Environmental Measurement and Modeling, Office of Research and Development, U.S. Environmental Protection Agency, Cincinnati, OH USA.

https://doi.org/10.1016/j.jenvman.2020.111506
Received 15 May 2020; Received in revised form 8 September 2020; Accepted 11 October 2020

Supervision, Project administration, Funding acquisition. Yu-Chen Wang: Methodology, Visualization, Writing - review & editing. Dale Robertson: Methodology, Formal analysis, Investigation, Writing - review & editing, Visualization. Donald Scavia: Writing - review & editing, Supervision, Project administration, Funding acquisition. Jay Martin: Writing - review & editing, Supervision, Project administration, Funding acquisition. Noel Aloysius: Methodology, Writing - review & editing. Anna Apostel: Methodology, Formal analysis, Writing - review & editing. Chelsie Boles: Methodology, Writing - review & editing. Michael Brooker: Formal analysis, Writing - review & editing. Remegio Confesor Methodology, Writing - review & editing. Awoke Teshager Dagnew: Methodology, Writing - review & editing. Tian Guo: Methodology, Writing - review & editing. Jeffrey Kast: Methodology, Writing - review & editing. Haley Kujawa: Methodology, Writing - review & editing. Asmita Murumkar: Methodology, Writing - review & editing. Rebecca Logsdon Muenich: Methodology, Writing - review & editing. Todd Redder: Methodology, Writing - review & editing.

1. Introduction

Anthropogenically enhanced nutrient loading has accelerated eutrophication in aquatic ecosystems around the world (Anderson et al., 2002). Affected ecosystems may develop harmful algal blooms (HABs) or hypoxia that disrupt natural systems and adversely affect human health and economic well-being (Hoagland et al., 2002). Conservation and restoration efforts aimed at mitigating nutrient loadings often seek to identify a watershed's critical source areas (CSAs; alternatively 'hotspots' or 'priority management areas'), defined as areas with disproportionately elevated nutrient and sediment losses (White et al., 2009). Conservation and restoration actions can then be prioritized or 'targeted' at these locations to ensure efficient use of limited resources.

Watershed-scale hydrologic models are one of several tools available to facilitate CSA identification. These models can simulate nutrient and sediment losses across a landscape and be used to identify CSA locations (e.g., Tripathi et al., 2003; Srinivasan et al., 2005; Strauss et al., 2007; Robertson et al., 2009;White et al., 2009; Georgas et al., 2009; Ghebremichael et al., 2010, 2013; Kovacs et al., 2012; Wang and Cui 2012; Robertson and Saad, 2019). Watershed models are particularly helpful in identifying CSAs across large spatial scales or in data-poor environments. However, as we show here, these simulations are affected by uncertainty.

Model uncertainty may arise from differing, but equally valid, model representations driven by differences in (1) input data (e.g., multiple data sets may characterize the rate of fertilizer application differently across a watershed), (2) parameterization (e.g., different values may be assigned to model parameters), and (3) model structure or system conception (e.g., dominant processes may be described using different mathematical formulations) (Wagener and Gupta 2005). Where different model representations are valid, the differences among resultant predictions (i.e., the uncertainty) warrent consideration in policy and management decisions.

Earlier work has shown that CSA simulations from watershed-scale hydrologic models may vary due to alternative input data (Xu et al., 2016), parameterization (Robertson et al., 2009), and variations in model structure and system conception (Niraula et al., 2013; Wang et al., 2016a). Our goal was to provide a more comprehensive assessment of CSA uncertainty using an ensemble of five hydrologic models. Our objectives were to (1) simulate CSA locations, (2) quantify CSA uncertainty expressed as variation across ensemble models, and (3) determine the extent to which CSA uncertainty was affected by model differences in fertilizer application rates, which we hypothesized would substantially affect CSA simulation uncertainty.

2. Study area

We focused on the ~17,000-km^2 Maumee River watershed, the largest tributary of Lake Erie – one of the largest freshwater lakes in the world. The watershed is dominated by row crop agriculture (~70%) and is a primary source of phosphorus (P) loadings that contribute to the occurrence of HABs (Obenour et al., 2014; Bertani et al., 2016; Maccoux et al., 2016; Stumpf et al., 2016; Verhamme et al., 2016; Manning et al., 2019). Hence, much effort has been devoted to evaluating the capacity of agricultural management actions (or 'best management practices'; BMPs) to reduce P export from the Maumee River watershed's agricultural fields to meet recently established loading goals (Scavia et al., 2017; Martin et al., 2019, and Martin et al. *In Review*). Identification of CSAs would aid the placement of BMPs in the Maumee River watershed.

3. Methods

3.1. The model ensemble

We used an ensemble of five watershed-scale models that were constructed by different research groups that made different choices with respect to model inputs, parameterization, and structure. Each model constituted an equally defensible representation of the watershed, which resulted in a range of simulations reflecting CSA uncertainty. The ensemble was composed of four Soil and Water Assessment Tool (SWAT) models and one SPAtially Referenced Regression On Watershed attributes (SPARROW) model. The SWAT model is a mechanistic hydrologic model that simulates hydrological and biogeochemical processes including, but not limited to, infiltration-runoff, evapotranspiration, and biogeochemical processing and transport (Arnold et al., 2012). The SPARROW model is a hybrid statistical-mechanistic model that uses regression-based coefficients to describe relationships between observed nutrient loadings and watershed characteristics (Smith et al., 1997). Both SWAT and SPARROW have been used extensively to identify CSAs (for SWAT, see White et al., 2009, Ghebremichael et al., 2010, 2013, Niraula et al., 2011, 2013, Liu et al., 2016, Xu et al., 2016, Teshager et al., 2017; for SPARROW, see Brakebill et al., 2010, Robertson et al., 2009, 2014; Robertson and Saad, 2011).

The five models were built and tested by separate research groups; HU = Heidelberg University (SWAT), LT = LimnoTech Inc. (SWAT), UM = University of Michigan (SWAT), OSU = Ohio State University (SWAT), GS = U.S. Geological Survey (SPARROW). The models varied with respect to their *input data* and associated assumptions depicting an array of natural and anthropogenic processes including, for example, fertilizer application rates across the watershed. The SWAT models also varied with respect to *parameterization* because they used different parameter values when calibrating to observed daily streamflow and nutrient and sediment loads at the gauge near the watershed outlet (U.S. Geological Survey (USGS) streamgage #04193500; U.S. Geological Survey, 2020). The GS SPARROW model was calibrated to annual streamflow and nutrient loads from >1100 gauged locations throughout the Upper Midwest of the United States (Robertson and Saad, 2019). The five models varied with respect to *structure* because the SWAT models used different sub-modules and equations to simulate specific processes, and the SPARROW model depicted the system via an entirely different set of mathematical formulations.

We included two versions of each model used in the ensemble - differing with respect to model input data, spatial discretization, and parameterization - to assess how CSA simulations changed due to model improvements. The models in version 1 (v.1) of the ensemble were those previously reported by Scavia et al. (2017) and Robertson and Saad (2011) while the models in version 2 (v.2) of the ensemble were those previously reported by Martin et al. (2019), Martin et al. (*In Review)* and Robertson and Saad (2019) . The v.1 and v.2 SWAT models differed with respect to the input data and assumptions used to describe tillage, buffer strips, tile drains, and cover crop practices. In v.1 SWAT models, the representation of these practices was determined without consultation across modeling groups while their representation in the v.2 SWAT

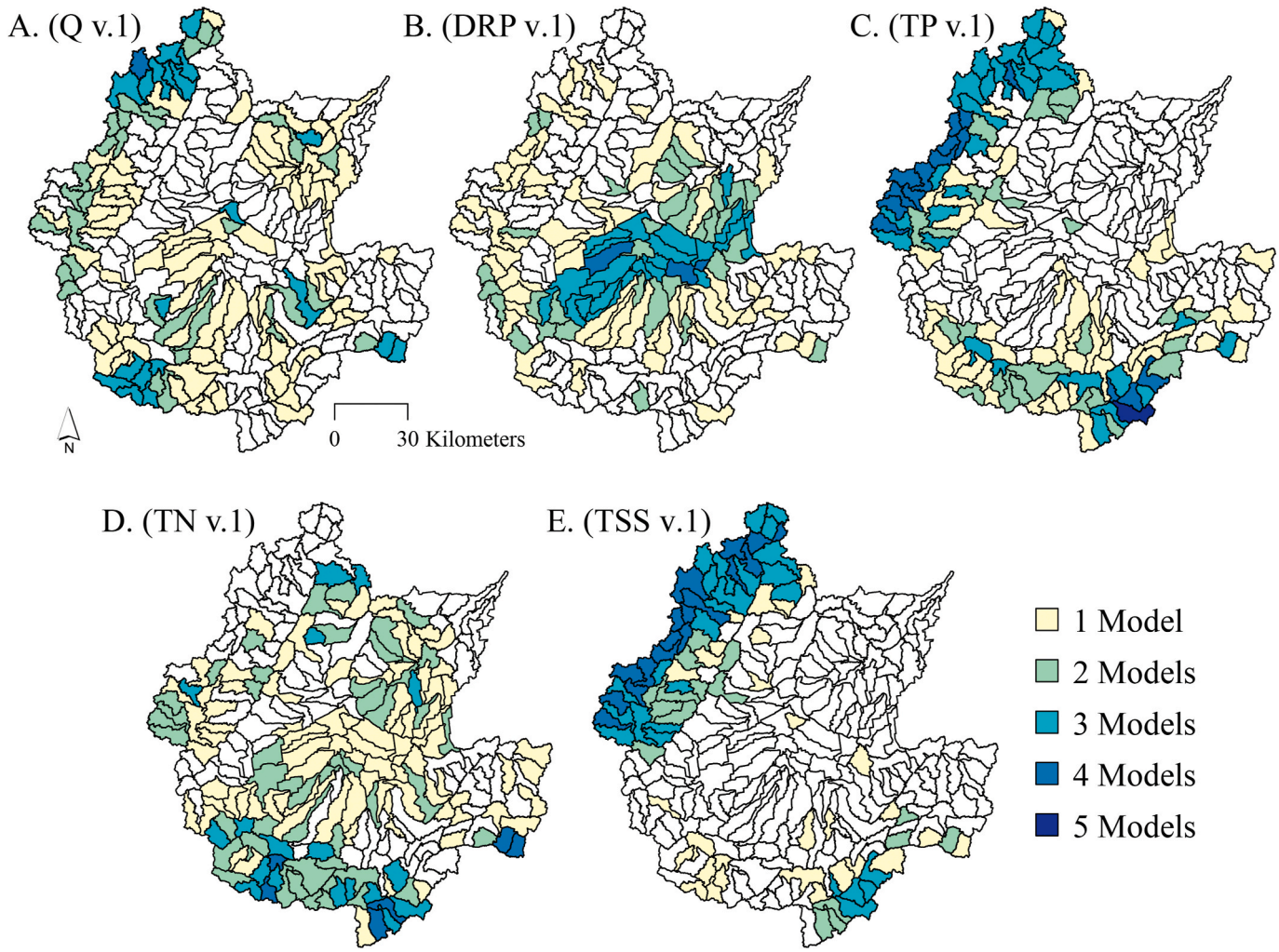

Fig. 1. Summary critical source area (CSA) maps for version 1 (v.1) of the model ensemble for the Maumee River watershed. Each HUC12 watershed was assigned a value (0–5), indicating the number of ensemble models that identified the watershed as a CSA. Version 1 of the GS model did not provide predictions for Q, DRP, or TSS. Hence, the maximum number of models for subfigures A, B, and E was 4. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

models was jointly decided upon via analysis of regional survey data. The v.1 and v.2 SWAT models also differed with respect to their representation of the magnitude and location of fertilizer applications. The v.1 SWAT models applied fertilizer by varying methods, including agronomic rates (i.e. rates sufficient to sustain crop growth) and whole-watershed mass balance approaches, while the v.2 SWAT models determined application rates via county-level fertilizer sales and livestock estimates, an approach similar to that used by the v.1 and v.2 GS SPARROW models. The v.1 and v.2 GS SPARROW models differed with respect to the spatial resolution of the individual catchments and the statistical techniques used to evaluate model performance across calibration sites, which resulted in improved accuracy for the v.2 model. Model performance statistics are listed in the Supplemental Material (Table S1), but see Robertson and Saad (2011, 2019), Scavia et al. (2017), and Martin et al. (*In Review)* for additional discussion.

3.2. Critical source areas identification

We identified CSAs in three steps. First, we extracted each model's mean annual subwatershed discharge (Q), and export of dissolved reactive phosphorus (DRP), total phosphorus (TP), total nitrogen (TN), and total suspended solids (TSS). Note that v.1 of the GS model did not provide Q, DRP, or TSS, and v.2 of the GS model did not provide DRP. These mean annual export values were calculated over three time-periods: 2005–2014 for all of the SWAT models (v.1 and v.2), 1971–2006 for the v.1 SPARROW model, and 2000–2014 for the v.2 SPARROW model. Model time-periods varied because the SWAT models necessitated a 5-year warm-up period (2000–2004) to stabilize initial conditions, and the v.1 SPARROW model was originally built for a different research effort. Second, because the models used subwatersheds of varying size, area-weighted averages were used to convert individual model outputs to USGS Hydrologic Unit Code 12 (HUC-12) subwatersheds (Seaber et al., 1987). Third, CSAs were identified using each model, representing the 20% of HUC-12 subwatersheds with the highest loss rates (White et al., 2009; Wang et al., 2016b).

3.3. Critical source area model agreement

We assessed CSA agreement in three ways. First, we counted the number of models that identified each HUC-12 subwatershed as a CSA based on maps for each model output variable (Q, DRP, TP, TN, and TSS) in each version of the ensemble (v.1, v.2). Second, we calculated the mean percent agreement ($\overline{PA}$) between individual models for each model output variable (Q, DRP, TP, TN, and TSS) in each version of the ensemble (v.1, v.2) as,

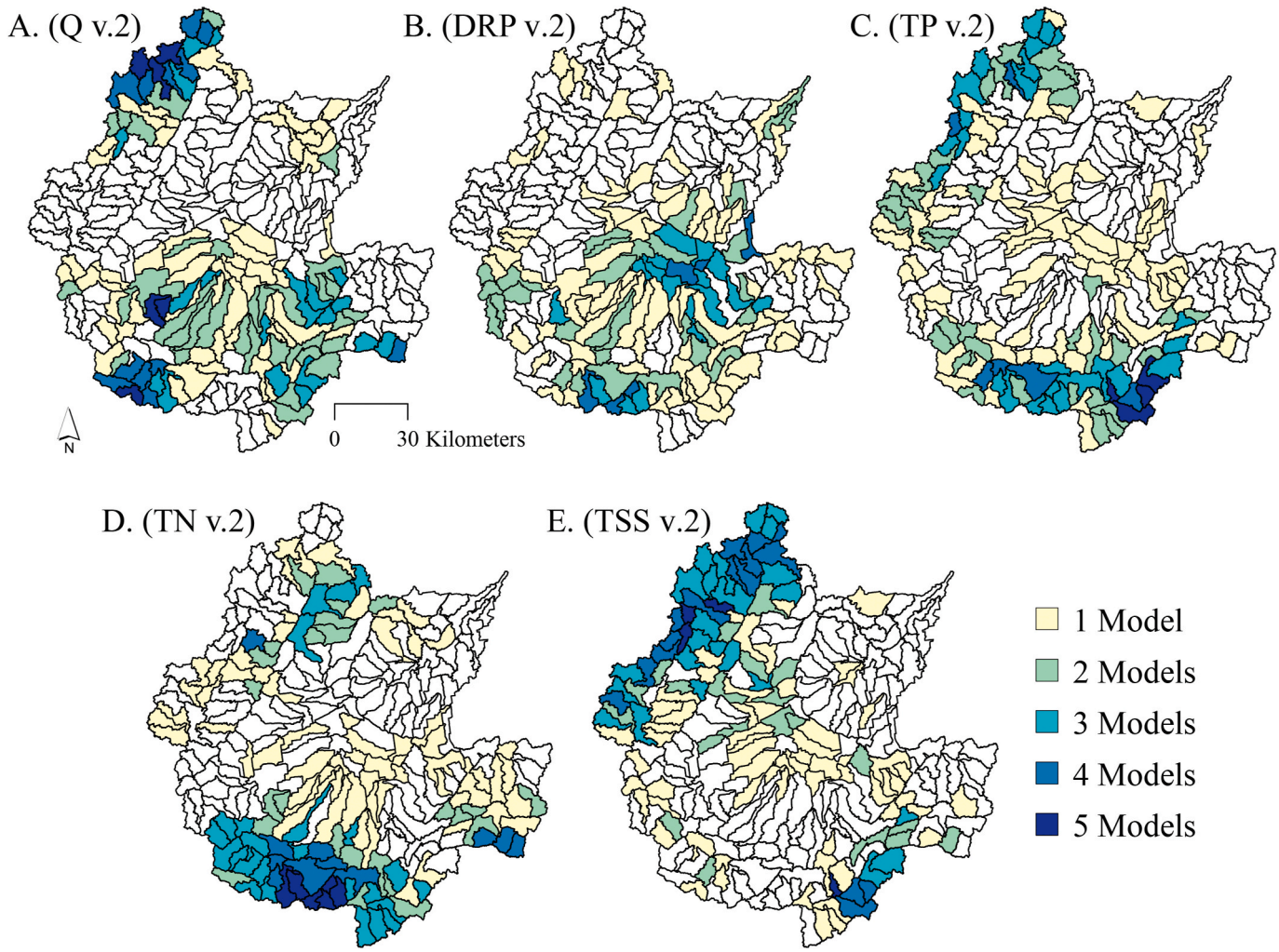

Fig. 2. Summary critical source area (CSA) maps for version 2 of the model ensemble. Each HUC12 watershed was assigned a value (0–5), indicating the number of ensemble models that identified the watershed as a CSA. Version 2 of the GS model did not provide predictions for DRP. Hence, the maximum number of models for subfigure B was 4. The maximum number of models for subfigure B was 4. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

$$\overline{PA} = \frac{\sum_i^n (x_i/y_i)}{n} \times 100 \quad (1)$$

where $\overline{PA}$ is the mean percent agreement, n is the total number of possible model pairings in the ensemble ($n = 6$ or 10, depending on the model output variable and ensemble version), i is an index for the model pairings, x_i is the number of HUC-12 watersheds that were designated as a CSA by both models in the pairing, and y_i is the number of HUC-12 watersheds that were designated as a CSA by either model in the pairing. Third, we calculated the mean Spearman rank-order correlation coefficient ($\overline{r_s}$) between individual models for each model output variable (Q, DRP, TP, TN, and TSS) in each version of the ensemble (v.1, v.2). We used the r_s statistic to assess correlation between the model's rankings (1st to 252nd, the total number of subwatersheds) of HUC-12 subwatersheds with respect to model output variables. The r_s statistic ranges from −1 to 1, where −1 indicates perfect negative correlation and 1 indicates perfect positive correlation. We calculated $\overline{r_s}$ as,

$$\overline{r_s} = \frac{\sum_i^n r_{s,i}}{n} \times 100 \quad (2)$$

where $\overline{r_s}$ is the mean Spearman rank-order correlation coefficient (−1 to 1), n is the total number of possible model pairings in the ensemble ($n =$ 6 or 10, depending on the model output variable and ensemble version), i is an index for the model pairings, $r_{s,i}$ is the Spearman rank-order correlation coefficient calculated between the HUC-12 rankings for the models in the pairing.

3.4. Fertilizer application effects on nutrient CSA location

Because fertilizer applications are a primary source of nutrient input in agricultural landscapes, we hypothesized that simulated CSAs would correspond to locations where more fertilizer was applied, and that CSA uncertainty would be partially explained by differences in how the models represented the distribution of fertilizer across the watershed. We tested this by assessing the coefficient of determination (R^2) between the quantity of fertilizer applied (sum of mineral and organic P or N) and simulated nutrient export (DRP, TP, or TN) for each HUC-12 subwatershed. Further, we assessed the extent to which the models differed in their representation of fertilizer application in the watershed by calculating the correlation coefficient (r) between model-by-model representations of the quantity of fertilizer applied (sum of mineral and organic P or N) across each HUC-12 subwatershed. We limited this part of the analysis to v.2 of the ensemble because fertilizer inputs from all v.1 models were not available.

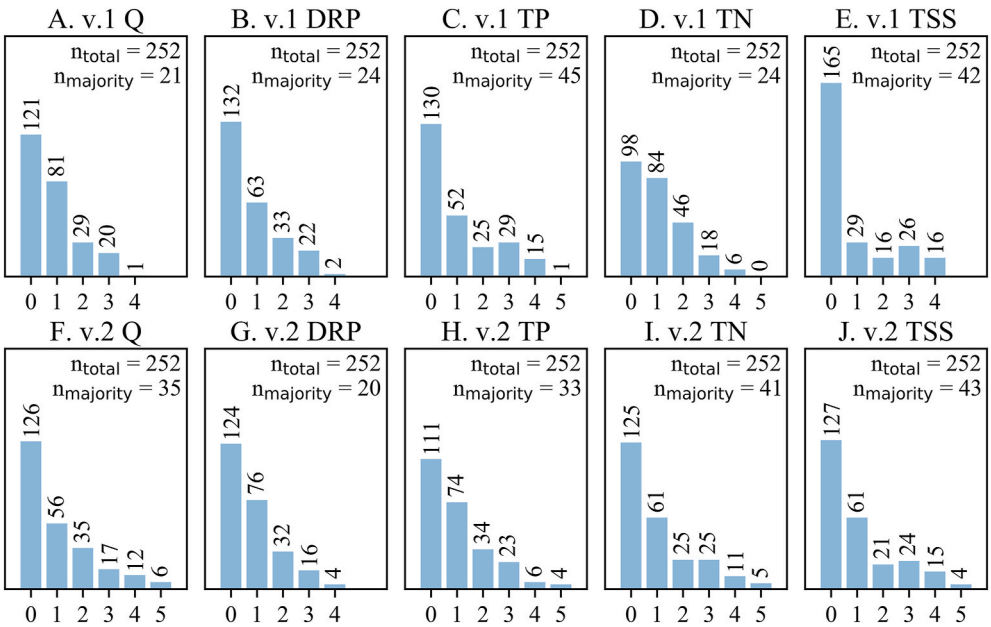

Fig. 3. (Color): Histograms of the number of HUC-12s with each model count value (0–5) (see Figs. 1–2) for each model output variable (DRP, Q, TN, TP, or TSS) and ensemble version (v.1 or v.2). n_{total} = the number of HUC-12s that were identified as a CSA by at least one of the models; $n_{majority}$ = the number of HUC-12s that were identified as a CSA by a majority of models (≥ 3 models). The total number of HUC-12s in the Maumee River watershed was 252. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 1
The mean percent agreement ($\overline{PA}$) and mean spearman rank-order correlation ($\overline{r_s}$) statistics were calculated across all possible model pairings in the ensemble, for each model output variable (Q, DRP, TP, TN, TSS) in each ensemble version (v.1, v.2). The *PA* and r_s statistics for individual model pairing are provided in the Supplemental Material (Tables S1-S2).

	$\overline{PA}$		$\overline{r_s}$	
	v.1	v.2	v.1	v.2
Q	20%	28%	0.38	0.57
DRP	23%	21%	0.34	0.37
TP	28%	22%	0.33	0.33
TN	16%	27%	0.21	0.39
TSS	46%	30%	0.56	0.40

Table 2
Pearson correlation coefficients (*r*) measuring correlation in fertilizer application across HUC-12 subwatersheds. The statistics were calculated for each possible model pairing, for total P and N applied (sum of mineral and organic P or N, respectively).

		P Applied				N Applied			
		LT	UM	OSU	GS	LT	UM	OSU	GS
v.2	**HU**	0.42	0.63	0.66	0.73	0.38	0.68	0.45	0.66
	LT		0.49	0.50	0.49		0.50	0.43	0.46
	UM			0.65	0.80			0.39	0.77
	OSU				0.76				0.49

4. Results

4.1. Critical source areas agreement

The models agreed regarding the location of a subset of CSAs (Figs. 1–2). For example, TP and TSS CSAs for v.1 appeared to coalesce along the northwestern boundary of the watershed (Fig. 1C and E), and DRP (Fig. 1B) along the Maumee River. In addition, there were 20–45 HUC-12s identified as CSAs by 3 or more models across all output variables and ensemble versions (see $n_{majority}$ values in Fig. 3).

However, the ensembles more often disagreed regarding CSA location. The overwhelming majority of HUC-12s identified as CSAs were identified as such by a minority of models (see the generally skewed distribution towards lower model counts in Fig. 3). The $\overline{PA}$ statistics indicated that, on average, only 16–46% of HUC-12 subwatersheds simulated as a CSA by one model, were also simulated as a CSA in another model (Table 1). Further, the $\overline{r_s}$ statistics indicated that although the models' overall ranking of subwatersheds was positively correlated, the strength of the relationship was generally 'weak' to 'moderate' ($\overline{r_s}$ values ranged from 0.21 to 0.57, Tables 1, S1, S2) (Gordon et al., 2004).

4.2. Fertilizer application impacts on nutrient CSA location

We hypothesized that CSA simulations may differ because ensemble models applied fertilizer across the watershed at different rates (Table 2). We observed weak correlation between some model pairing with respect to the quantity of fertilizer applied across HUC-12 subwatersheds. The lack of correlation was especially apparent with regard to the LT model, which appeared as a relative outlier with *r*-values ranging from 0.38 to 0.50 (Table 2). Yet even where the models did apply fertilizer at similar rates across the HUC-12 subwatersheds, fertilizer application did not appear to be a strong determinant of nutrient export – and therefore CSA location – across all models (Fig. 4). Fertilizer application rate was powerful in explaining nutrient loads within the GS model (see R^2 values of 0.74 and 0.76 for TN and TP, respectively; and this would be expected because the GS model is based on regression relationships). However, fertilizer application rates were only weakly related to nutrient export and thus CSA location for most other models. For example, fertilizer application rates explained less than 4% of the variation in DRP, TP, and TN export from the OSU model.

5. Discussion

5.1. CSA location simulation uncertainty

There was generally weak agreement among models regarding CSA locations. Our results were consistent with Niraula et al. (2013), who showed that CSA locations varied between model types (SWAT and the Generalized Watershed Loading Function model); with Xu et al. (2016), who showed that CSA locations differed across variations of the SWAT model using digital elevation models (DEMs) of differing sources; and with Robertson et al. (2009) who showed that CSA locations varied by model parameterization.

We found more pronounced variation across models than did Niraula

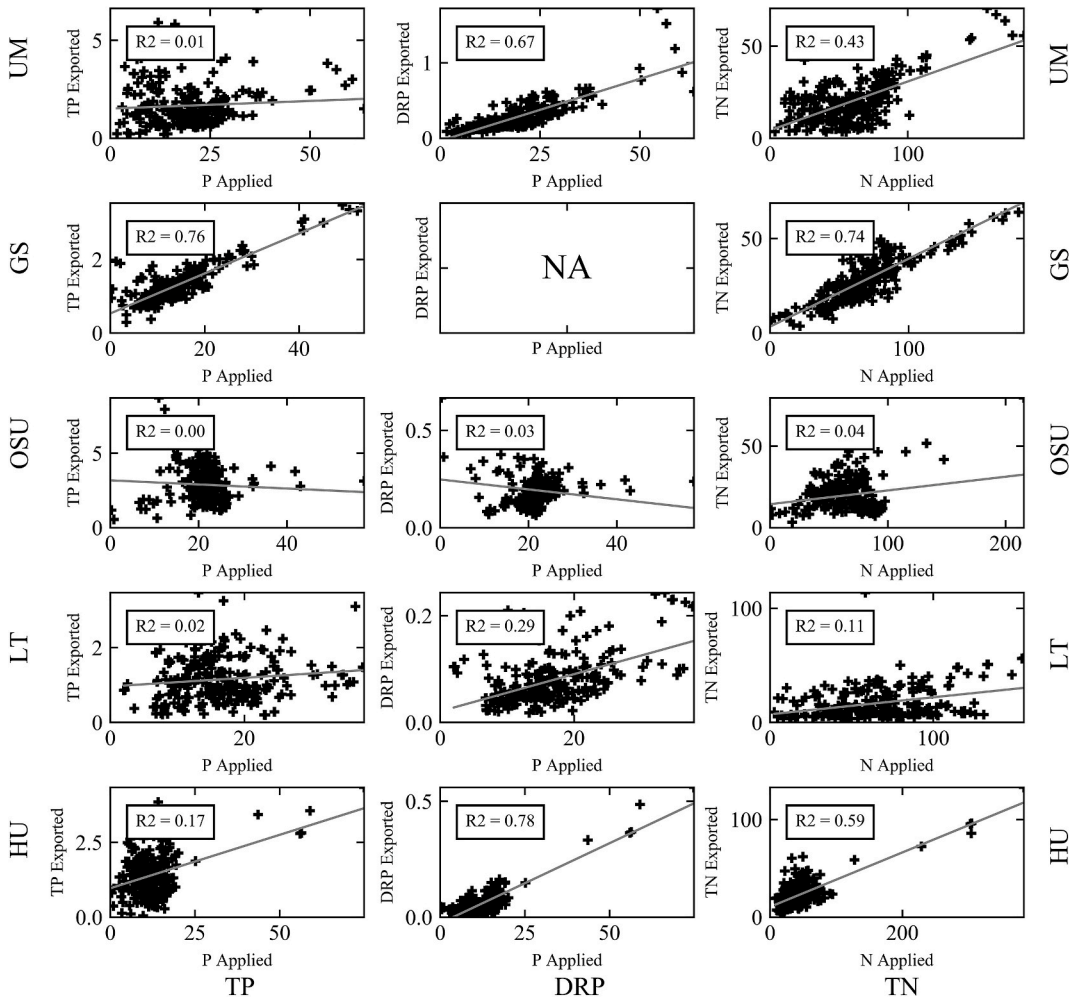

Fig. 4. Scatterplots showing the relationship between predicted nutrient export (TP, DRP, or TN) and total fertilizer applied (sum of mineral and organic N or P). Each marker corresponds to a HUC-12 subwatershed (n = 252). All axes are presented in kg ha^{-1} yr^{-1}. The scatterplots depict only v.2 models as fertilizer inputs from all v.1 models were not available.

et al. (2013). Our calculated *PA* values for their results were 83%, 64%, and 53% for TSS, TP, and TN, respectively, which were substantially higher than most values for our models (Table 1; and for the *PA* statistics calculated for individual model pairings, see Supplemental Material Table S2). This suggests that more varied model representation - as depicted by differences across our models - cause more uncertainty than when evaluating the effects of model inputs, parameterization, or structure alone. We believe that our assessment of CSA prediction uncertainty is conservative because our ensemble included just two types of models (SWAT and SPARROW). A more diverse ensemble of model-types could lead to further variation in CSA simulations.

5.2. What led to uncertainty in the CSA predictions?

We hypothesized that CSA uncertainty would be attributable to differences in the rate of fertilizer application across the HUC-12 subwatersheds. While our results confirm that models are more likely to predict CSAs in areas that receive more fertilizer (see generally positive relations in Fig. 4), we found that fertilizer application rates did not explain most variation in load predictions across models and output variables. For example, the UM and GS models applied P and N fertilizer at similar rates across the HUC-12 subwatersheds (*r*-statistic values of 0.80 and 0.77 for P and N fertilizer, respectively), but their CSA simulations diverged (*PA* values of 16% and 28% for TP and TN, respectively). Hence, we were unable to attribute CSA uncertainty to differences in how the models depicted fertilizer application across the watershed.

While we could not attribute CSA uncertainty to individual aspects of model variation (i.e. input data, parameterization, and model structure), our results suggest that structural differences may be particularly powerful in affecting model output uncertainty. The GS SPARROW model had consistently lower *PA* statistics relative to the SWAT models (see Supplemental Material Table S2), indicating that structural variation increased CSA disagreement. This indicates that future analyses would benefit from attention to potential differences in model structure as affecting CSA uncertainty. This finding also highlights the utility of ensemble-based analyses that incorporate models with varied structures.

Our results may have been influenced by the relatively large scale at which we assessed CSAs. We aggregated each model's simulations – often originating at relatively finer spatial scales - to HUC-12 subwatersheds. While necessary to facilitate inter-model comparisons, use of these relatively large spatial units may have resulted in 'information loss,' whereby simulated loads were homogenized via averaging (Wang et al., 2016b). The application of hydrologic models at alternative spatial scales may result in differing accounts of model agreement. Our results may also have been influenced by the lack of variability in land use in the intensively agricultural Maumee River watershed.

5.3. Are the CSA predictions useful?

Uncertainty in simulated CSA locations does not necessarily indicate that these simulations lack utility. We found that there was a subset of HUC-12s that a majority of ensemble models identified as CSAs (Fig. 3). These HUC-12s may be prioritized or 'targeted' for conservation and restoration action with relative confidence, demonstrating the utility of watershed-scale hydrologic models as applied to identify CSAs. Our results also indicate that these models may be helpful in identifying low-priority areas, i.e. areas that may be deferred from consideration of conservation and restoration resources (Robertson et al., 2009), as 98 to 165 (~39%–65%) of the 252 HUC-12s were not identified by any of the models, depending upon the model output variable and ensemble version.

Yet our results indictate that the application of models to identify CSA locations would benefit from comprehensive uncertainty analyses that evaluate all sources of model uncertainty (i.e., input data, parameterization, and model structure). Ensemble-based methods are beneficial because of their potential to concurrently assess the effect of these sources, as our work demonstrates. Managers and decision makers should consult multiple models from diverse research groups wherever possible. Where multiple models do not exist for the watershed of interest, there is potential for new methods of model development to evaluate diverse sources of model uncertainty, such as the Integrated Parameter Estimation and Uncertainty Analysis Tool (IPEAT, Yen et al., 2014). Further, a variety of methods may be applied to assess the effects of parameter uncertainty alone, such as Monte Carlo, Latin Hypercube or bootstrapping sampling, or more formalized uncertainty procedures such as Generalized Likelihood Uncertainty Estimation (GLUE) (Shirmohammadi et al., 2006; Robertson et al., 2009).

Finally, there are opportunities to reduce CSA prediction uncertainty. Uncertainty may be reduced via use of additional in-stream observations (Daggupati et al., 2015), spatially distributed and remotely-sensed evapotranspiration and soil moisture data (Houser et al., 1998; Chen et al., 2011; Herman et al., 2018; Rajib et al., 2018), spatially distributed data describing initial or legacy soil nutrient levels (Jarvie et al., 2013; Dayton et al., 2020), and additional or improved data describing agricultural management practices - for example, depicting the type, timing, and manor of fertilization and tillage practices (Muenich et al., 2017).

6. Conclusion

We found that a small subset of HUC-12 subwatersheds were simulated as CSAs by a majority of models and that these areas could be targeted for conservation and restoration actions with relative confidence. We also found that a large subset of subwatersheds were not simulated as a CSA by any of the models and that these areas could be removed for consideration for conservation and restoration action. However, we conclude that there is substantial uncertainty in CSA predictions provided by watershed-scale hydrologic models, as CSA location predictions may vary as provided by different models of the same watershed, especially if there are considerable differences with respect to model structure. Hence, models applied to predict CSA location would benefit from comprehensive uncertainty analyses.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

We wish to thank members of our stakeholder advisory group and the organizations they represented for providing guidance and participating in outreach activities. We would also like to thank Katherine Merriman with the U.S. Geological Survey New York Water Science Center and two anonymous reviewers of the Journal of Environmental Management for helpful comments on earlier drafts of this paper. This work was funded by the Ohio Department of Higher Education Harmful Algal Bloom Research Initiative (R/HAB-5-ODHE) through Ohio Sea Grant. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jenvman.2020.111506.

References

Anderson, D.M., Glibert, P.M., Burkholder, J.M., 2002. Harmful algal blooms and eutrophication: nutrient sources, composition, and consequences. Estuaries 25 (4), 704–726.

Arnold, J.G., Moriasi, D.N., Gassman, P.W., Abbaspour, K.C., White, M.J., Srinivasan, R., et al., 2012. SWAT: model use, calibration, and validation. Transactions of the ASABE 55 (4), 1491–1508.

Bertani, I., Obenour, D.R., Steger, C.E., Stow, C.A., Gronewold, A.D., Scavia, D., 2016. Probabilistically assessing the role of nutrient loading in harmful algal bloom formation in western Lake Erie. J. Great Lake. Res. 42 (6), 1184–1192.

Brakebill, J.W., Ator, S.W., Schwarz, G.E., 2010. Sources of suspended-sediment flux in streams of the chesapeake bay watershed: a regional application of the SPARROW model 1. Journal of the American Water Resources Association 46 (4), 757–776.

Chen, F., Crow, W.T., Starks, P.J., Moriasi, D.N., 2011. Improving hydrologic predictions of a catchment model via assimilation of surface soil moisture. Adv. Water Resour. 34 (4), 526–536.

Daggupati, P., Yen, H., White, M.J., Srinivasan, R., Arnold, J.G., Keitzer, C.S., Sowa, S.P., 2015. Impact of model development, calibration and validation decisions on hydrological simulations in West Lake Erie Basin. Hydrol. Process. 29 (26), 5307–5320.

Dayton, E.A., Shrestha, R.K., Fulford, A.M., Love, K.R., Culman, S.W., Lindsey, L.E., 2020. Soil test phosphorus and phosphorus balance trends: a county-level analysis in Ohio. Agron. J. https://doi.org/10.1002/agj2.20146.

Georgas, N., Rangarajan, S., Farley, K.J., Jagupilla, S.C.K., 2009. AVGWLF-Based estimation of nonpoint source nitrogen loads generated within long island sound subwatersheds 1. Journal of the American Water Resources Association 45 (3), 715–733.

Ghebremichael, L.T., Veith, T.L., Watzin, M.C., 2010. Determination of critical source areas for phosphorus loss: Lake Champlain basin, Vermont. Transactions of the ASABE 53 (5), 1595–1604.

Ghebremichael, L.T., Veith, T.L., Hamlett, J.M., 2013. Integrated watershed-and farm-scale modeling framework for targeting critical source areas while maintaining farm economic viability. J. Environ. Manag. 114, 381–394.

Gordon, N.D., McMahon, T.A., Finlayson, B.L., Gippel, C.J., Nathan, R.J., 2004. Stream Hydrology: an Introduction for Ecologists. John Wiley and Sons.

Herman, M.R., Nejadhashemi, A.P., Abouali, M., Hernandez-Suarez, J.S., Daneshvar, F., Zhang, Z., et al., 2018. Evaluating the role of evapotranspiration remote sensing data in improving hydrological modeling predictability. J. Hydrol. 556, 39–49.

Hoagland, P.A.D.M., Anderson, D.M., Kaoru, Y., White, A.W., 2002. The economic effects of harmful algal blooms in the United States: estimates, assessment issues, and information needs. Estuaries 25 (4), 819–837.

Houser, P.R., Shuttleworth, W.J., Famiglietti, J.S., Gupta, H.V., Syed, K.H., Goodrich, D. C., 1998. Integration of soil moisture remote sensing and hydrologic modeling using data assimilation. Water Resour. Res. 34 (12), 3405–3420.

Jarvie, H.P., Sharpley, A.N., Spears, B., Buda, A.R., May, L., Kleinman, P.J., 2013. Water Quality Remediation Faces Unprecedented Challenges from "Legacy Phosphorus".

Kovacs, A., Honti, M., Zessner, M., Eder, A., Clement, A., Blöschl, G., 2012. Identification of phosphorus emission hotspots in agricultural catchments. Sci. Total Environ. 433, 74–88.

Liu, R., Xu, F., Zhang, P., Yu, W., Men, C., 2016. Identifying non-point source critical source areas based on multi-factors at a basin scale with SWAT. J. Hydrol. 533, 379–388.

Maccoux, M.J., Dove, A., Backus, S.M., Dolan, D.M., 2016. Total and soluble reactive phosphorus loadings to Lake Erie: a detailed accounting by year, basin, country, and tributary. J. Great Lake. Res. 42 (6), 1151–1165.

Manning, N.F., Wang, Y.C., Long, C.M., Bertani, I., Sayers, M.J., Bosse, K.R., et al., 2019. Extending the forecast model: predicting Western Lake Erie harmful algal blooms at multiple spatial scales. J. Great Lake. Res. 45 (3), 587–595.

Martin, J.F., Kalcic, M.M., Aloysius, N., Apostel, A.M., Brooker, M.R., Evenson, G.R., Kast, J.B., Kujawa, H., Murumkar, A., Becker, R., Boles, C., Redder, T., Confesor, R., Guo, T., Dagnew, A., Long, C.M., Muenich, R., Scavia, D., Wang, Y., Robertson, D., 2019. Evaluating Management Options to Reduce Lake Erie Algal Blooms with Models of the Maumee River Watershed.Final Project Report - OSU Knowledge Exchange. Available at. http://kx.osu.edu/project/environment/habri-multi-model.

Martin, J.F., Kalcic, M.M., Aloysius, N., Apostel, A.M., Brooker, M.R., Evenson, G.R., Kast, J.B., Kujawa, H., Murumkar, A., Becker, R., Boles, C., Redder, T., Confesor, R. , Guo, T., Dagnew, A., Long, C.M., Muenich, R., Scavia, D., Wang, Y., Robertson, D. (In review) evaluating management options to reduce Lake Erie algal blooms using an ensemble of watershed models. J. Environ. Manag..

Muenich, R.L., Kalcic, M.M., Winsten, J., Fisher, K., Day, M., O'Neil, G., et al., 2017. Pay-for-performance conservation using SWAT highlights need for field-level agricultural conservation. Transactions of the ASABE 60 (6), 1925–1937.

Niraula, R., Kalin, L., Srivastava, P., Anderson, C.J., 2013. Identifying critical source areas of nonpoint source pollution with SWAT and GWLF. Ecol. Model. 268, 123–133.

Niraula, R., Kalin, L., Wang, R., Srivastava, P., 2011. Determining nutrient and sediment critical source areas with SWAT: effect of lumped calibration. Transactions of the ASABE 55 (1), 137–147.

Obenour, D.R., Gronewold, A.D., Stow, C.A., Scavia, D., 2014. Using a Bayesian hierarchical model to improve Lake Erie cyanobacteria bloom forecasts. Water Resour. Res. 50 (10), 7847–7860.

Rajib, A., Evenson, G.R., Golden, H.E., Lane, C.R., 2018. Hydrologic model predictability improves with spatially explicit calibration using remotely sensed evapotranspiration and biophysical parameters. J. Hydrol. 567, 668–683.

Robertson, D.M., Schwarz, G.E., Saad, D.A., Alexander, R.A., 2009. Incorporating uncertainty into the ranking of SPARROW model nutrient yields from Mississippi/Atchafalaya River Basin watersheds. J. Am. Water Resour. Assoc. 45, 534–549. DOI: 10.1111/j.1752-1688.2009.00310.x.

Robertson, D.M., Saad, D.A., 2011. Nutrient inputs to the Laurentian Great Lakes by source and watershed estimated using SPARROW watershed models 1. JAWRA Journal of the American Water Resources Association 47 (5), 1011–1033.

Robertson, D.M., Saad, D.A., 2019. Spatially referenced models of streamflow and nitrogen, phosphorus, and suspended sediment loads in streams of the Midwestern United States. https://doi.org/10.3133/sir20195114. U.S. Geological Survey Scientific Report 2019-5114.

Robertson, D.M., Saad, D.A., Schwarz, G.E., 2014. Spatial variability in nutrient transport by HUC 8, state, and subbasin based on Mississippi/atchafalaya River Basin SPARROW models. J. Am. Water Resour. Assoc. 50 (4), 988–1009.

Seaber, P.R., Kapinos, F.P., Knapp, G.L., 1987. Hydrologic Unit Maps: U.S. Geological Survey Water-Supply Paper 2294, p. 63. http://www.water.usgs.gov/GIS/huc.html.

Scavia, D., Kalcic, M., Muenich, R.L., Read, J., Aloysius, N., Bertani, I., et al., 2017. Multiple models guide strategies for agricultural nutrient reductions. Front. Ecol. Environ. 15 (3), 126–132.

Shirmohammadi, A., Chaubey, I., Harmel, R.D., Bosch, D.D., Muñoz-Carpena, R., Dharmasri, C., et al., 2006. Uncertainty in TMDL models. Transactions of the ASABE 49 (4), 1033–1049.

Smith, R.A., Schwarz, G.E., Alexander, R.B., 1997. Regional interpretation of water-quality monitoring data. Water Resour. Res. 33 (12), 2781–2798.

Srinivasan, M.S., Gérard-Marchant, P., Veith, T.L., Gburek, W.J., Steenhuis, T.S., 2005. Watershed scale modeling of critical sources areas of runoff generation and phosphorus transport. J. Am. Water Resour. Assoc. 41 (2), 361–377.

Strauss, P., Leone, A., Ripa, M.N., Turpin, N., Lescot, J.M., Laplana, R., 2007. Using critical source areas for targeting cost-effective best management practices to mitigate phosphorus and sediment transfer at the watershed scale. Soil Use Manag. 23, 144–153.

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

Teshager, A.D., Gassman, P.W., Secchi, S., Schoof, J.T., 2017. Simulation of targeted pollutant-mitigation-strategies to reduce nitrate and sediment hotspots in agricultural watershed. Sci. Total Environ. 607, 1188–1200.

Tripathi, M.P., Panda, R.K., Raghuwanshi, N.S., 2003. Identification and prioritisation of critical sub-watersheds for soil conservation management using the SWAT model. Biosyst. Eng. 85 (3), 365–379.

U.S. Geological Survey, 2020. USGS water data for the Nation. U.S. Geological Survey National Water Information System database. https://doi.org/10.5066/F7P55KJN. (Accessed 5 October 2020).

Verhamme, E.M., Redder, T.M., Schlea, D.A., Grush, J., Bratton, J.F., DePinto, J.V., 2016. Development of the western Lake Erie ecosystem model (WLEEM): application to connect phosphorus loads to cyanobacteria biomass. J. Great Lake. Res. 42 (6), 1193–1205.

Wagener, T., Gupta, H.V., 2005. Model identification for hydrological forecasting under uncertainty. Stoch. Environ. Res. Risk Assess. 19 (6), 378–387.

Wang, X., Cui, P., 2012. Support soil conservation practices by identifying critical erosion areas within an American watershed using the GIS-AGNPS model. J. Spatial Hydrol. 5 (2).

Wang, Y., Montas, H.J., Brubaker, K.L., Leisnham, P.T., Shirmohammadi, A., Chanse, V., Rockler, A.K., 2016a. Impact of spatial discretization of hydrologic models on spatial distribution of nonpoint source pollution hotspots. J. Hydrol. Eng. 21 (12), 04016047.

Wang, G., Chen, L., Huang, Q., Xiao, Y., Shen, Z., 2016b. The influence of watershed subdivision level on model assessment and identification of non-point source priority management areas. Ecol. Eng. 87, 110–119.

White, M.J., Storm, D.E., Busteed, P.R., Stoodley, S.H., Phillips, S.J., 2009. Evaluating nonpoint source critical source area contributions at the watershed scale. J. Environ. Qual. 38 (4), 1654–1663.

Xu, F., Dong, G., Wang, Q., Liu, L., Yu, W., Men, C., Liu, R., 2016. Impacts of DEM uncertainties on critical source areas identification for non-point source pollution control based on SWAT model. J. Hydrol. 540, 355–367.

Yen, H., Wang, X., Fontane, D.G., Harmel, R.D., Arabi, M., 2014. A framework for propagation of uncertainty contributed by parameterization, input data, model structure, and calibration/validation data in watershed modeling. Environ. Model. Software 54, 211–221.