Integrating Water Quality Data with a Bayesian Network Model to Improve Spatial and Temporal Phosphorus Attribution: Application to the Maumee River Basin

Zihan Wei[1], Sarfaraz Alam[1], Miki Verma[1], Margaret Hilderbran[1], Yuchen Wu[1], Brandon Anderson[1], Daniel E Ho[1], and Jenny Suckale[1]

[1]Stanford University

December 7, 2022

Abstract

Surface water nutrient pollution, the primary cause of eutrophication, remains a major environmental concern in Western Lake Erie despite intergovernmental efforts to regulate nutrient sources. The Maumee River Basin has been the largest nutrient contributor. The two primary nutrients sources are inorganic fertilizer and livestock manure applied to croplands, which are later carried to the streams via runoff and soil erosion. Prior studies on nutrient source attribution have focused on large watersheds or counties at long time scales. Source attribution at finer spatiotemporal scales, which enables more effective nutrient management, remains a substantial challenge. This study aims to address this challenge by developing a portable network model framework for phosphorus source attribution at the subwatershed (HUC-12) scale. Since phosphorus release is uncertain, we combine excess phosphorus derived from manure and fertilizer application and crop uptake data, flow dynamics simulated by the SWAT model, and in-stream water quality measurements into a probabilistic framework and apply Approximate Bayesian Computation to attribute phosphorus contributions from subwatersheds. Our results show significant variability in subwatershed-scale phosphorus release that is lost in coarse-scale attribution. Phosphorus contributions attributed to the subwatersheds are on average lower than the excess phosphorus estimated by the nutrient balance approach adopted by environmental agencies. Phosphorus release is higher during spring planting than the growing period, with manure contributing more than inorganic fertilizer. By enabling source attribution at high spatiotemporal resolution, our lightweight and portable model framework is suitable for broad applications in environmental regulation and enforcement for other regions and pollutants.

Integrating Water Quality Data with a Bayesian Network Model to Improve Spatial and Temporal Phosphorus Attribution: Application to the Maumee River Basin

Zihan Wei[1*], Sarfaraz Alam[1,2*], Miki Verma[3*], Margaret Hilderbran[2], Yuchen Wu[4], Brandon Anderson[2], Daniel E. Ho[2†], Jenny Suckale[1†]

[*]Joint First Authors
[†]Equal co-supervision
[1]Department of Geophysics, Stanford University, Stanford, CA, USA
[2]Regulation, Evaluation, and Governance Lab, Stanford University, Stanford, CA, USA
[3]Symbolic Systems Program, Stanford University, Stanford, CA, USA
[4]Department of Statistics, Stanford University, Stanford, CA, USA

Key Points:

- Our lightweight and portable framework integrates input, water quality, and flow dynamics to attribute sources of surface-water phosphorus
- Phosphorus release varies significantly at the subwatershed scale and peaks in the spring planting period
- Manure contributes more to phosphorus release than inorganic fertilizer in the Maumee River Basin

Corresponding author: Zihan Wei, zihanwei@stanford.edu

Abstract

Surface water nutrient pollution, the primary cause of eutrophication, remains a major environmental concern in Western Lake Erie despite intergovernmental efforts to regulate nutrient sources. The Maumee River Basin has been the largest nutrient contributor. The two primary nutrients sources are inorganic fertilizer and livestock manure applied to croplands, which are later carried to the streams via runoff and soil erosion. Prior studies on nutrient source attribution have focused on large watersheds or counties at long time scales. Source attribution at finer spatiotemporal scales, which enables more effective nutrient management, remains a substantial challenge. This study aims to address this challenge by developing a portable network model framework for phosphorus source attribution at the subwatershed (HUC-12) scale. Since phosphorus release is uncertain, we combine excess phosphorus derived from manure and fertilizer application and crop uptake data, flow dynamics simulated by the SWAT model, and in-stream water quality measurements into a probabilistic framework and apply Approximate Bayesian Computation to attribute phosphorus contributions from subwatersheds. Our results show significant variability in subwatershed-scale phosphorus release that is lost in coarse-scale attribution. Phosphorus contributions attributed to the subwatersheds are on average lower than the excess phosphorus estimated by the nutrient balance approach adopted by environmental agencies. Phosphorus release is higher during spring planting than the growing period, with manure contributing more than inorganic fertilizer. By enabling source attribution at high spatiotemporal resolution, our lightweight and portable model framework is suitable for broad applications in environmental regulation and enforcement for other regions and pollutants.

Plain Language Summary

Nutrient pollution and severe algal blooms remain major problems in western Lake Erie despite intergovernmental efforts to regulate sources in the U.S. and Canada. The Maumee River Basin has been the largest nutrient contributor to western Lake Erie. Historically, distributed agricultural areas dominated the nutrient contributions to the rivers, where sources include animal waste and inorganic fertilizer. Prior studies of nutrient source attribution have focused on large watersheds or counties at long time scales; source attribution at finer spatiotemporal scales, which can enable more effective nutrient management, remains a substantial challenge. Our study addresses this challenge by attributing phosphorus release at the subwatershed scale using a lightweight network model framework. Since phosphorus release is uncertain, we integrated water-quality measurements, excess phosphorus availability over land, and flow dynamics into a probabilistic framework to attribute phosphorus release to different sources. Our model reveals significant spatial and temporal variability in phosphorus release, which is averaged out in the coarse-scale attribution calculated using sparsely deployed water-quality monitors. Being able to identify such variability can greatly benefit targeted enforcement by enabling prioritization of regions, time periods, and source types with higher pollutant release.

1 Introduction

Despite tremendous expenditures and efforts devoted to cleanup and mitigation in recent decades, surface water pollution remains a major environmental concern (Howarth et al., 2000; Keiser & Shapiro, 2019; Downing et al., 2021). While pollution in urban areas has decreased alongside upgrades to wastewater treatment systems (Stets et al., 2020), water quality has hardly improved and even continues to degrade in agricultural areas (Stoddard et al., 2016; Stets et al., 2020). Because urban and rural water pollution come from fundamentally different sources, interventions to improve water quality in one setting are often ineffective in the other.

Pollution sources in urban areas are mainly point sources, such as wastewater treatment plants and factories, which release treated effluent to natural water bodies. These point sources are regulated by the National Pollutant Discharge Elimination System (NPDES) as part of the Clean Water Act since 1972 (USEPA, 2003). In contrast, pollution in agricultural areas comes primarily from unregulated nonpoint sources, namely the runoff from extensive agricultural lands (Baker, 1992; Parry, 1998; Carpenter et al., 1998; Ongley et al., 2010; Shen et al., 2012). The pollutants loaded in runoff, which are mainly nutrients including various forms of phosphorus and nitrogen for optimizing agricultural yields, originate from inorganic fertilizer sold commercially and manure collected from concentrated animal feeding operations (CAFOs) (Baker, 1992; Kumar et al., 2013).

Excessive application of manure and inorganic fertilizer can result in high nutrient loss in runoff from agricultural land (Higgs et al., 2000; Weil & Brady, 2017), leading to eutrophication followed by harmful algal blooms (EWG, 2022). Such nutrient losses in runoff are likely to worsen with more extreme storms and floods due to climate change, which intensify runoff and soil erosion (Ramos & Martínez-Casasnovas, 2006; Whitehead et al., 2009; Weil & Brady, 2017). While controlling the application rate to reduce nutrient loss is the obvious solution, it is only practicable by first identifying the relative contributions of inorganic fertilizer and manure, because agricultural nutrient management requires optimization rather than minimization as done for point sources. However, as both inorganic fertilizer and manure provide similar nutrients needed by crops (Culman et al., 2020; EWG, 2021), quantifying their relative contributions presents a further challenge in addition to difficulties associated with spatial attribution of nonpoint sources.

Detailed spatial attribution of nonpoint sources remains a highly underdetermined problem due to the lack of water-quality data with both high spatial and temporal resolutions (OC Interagency WQI Workgroup, 2017). Information about concentrated animal feeding operation (CAFO) manure production and inorganic fertilizer application can help constrain the overall contributions of various source types (Falcone, 2021) and locations (ELPC, 2014) but does not directly measure pollutants release into waterways. Release can vary due to runoff volume, amount of pollutant available on the surface, and soil properties (Sharpley, 1995, 1997; Hart et al., 2004). More frequent and spatially dense measurements of pollutant concentrations in waterways would certainly improve our ability to detect pollution, but better detection does not necessarily solve the attribution problem.

There is a fundamental difference between pollutant detection and attribution. Detection is the physical measurement of pollutants, identifying whether pollutants are present and, if so, in what amount. In contrast, attribution refers to the process of determining the sources of emerging pollutants and the relative contributions of sources. Attributing pollution to specific sources is more challenging than merely detecting it, because attribution requires not only pollutant concentration data, but also modeling of physical processes of surface water pollutant transport, as well as a framework that establishes the possible connection between sources and pollutants.

The goal of this paper is to advance the ability to attribute phosphorus release to different sources at the subwatershed scale by integrating water-quality observations, phosphorus input information, and hydrological modeling into a portable network model framework. Our subwatersheds are comparable to USGS HUC-12 (12-digit Hydrologic Unit Code) watersheds. Our lightweight and portable network model estimates how much phosphorus is released from different subwatersheds. The network model integrates available waterway phosphorus measurements with simulated flow dynamics in the stream network from the commonly used Soil and Water Assessment Tool (SWAT) hydrologic model (Arnold et al., 2012; Kast et al., 2019). Since the phosphorus release is uncertain, we combine the data and model outputs into a probabilistic framework and apply statistically robust Approximate Bayesian Computation (ABC) (Beaumont et al., 2002; Sunnåker et al., 2013) to estimate ranges of phosphorus release from subwatersheds. Through cross validation, we also quantify the information gain from different water quality monitors, which can potentially help planning for additional monitor locations for improved attribution in the future.

Most prior attempts to attribute phosphorus to nonpoint sources adopt deterministic hydrologic models, where the phosphorus release from a watershed is a function of flow dynamics, soil properties, land use, and phosphorus availability (Kast et al., 2019; Easton et al., 2007). Such models include SWAT (Arnold et al., 2012; Kast et al., 2019), USGS SPARROW (Schwarz et al., 2006), EPA Storm Water Management Model (SWMM) (Gironás et al., 2010), EPA Hydrologic Simulation Program-Fortran (HSPF) (Bicknell et al., 1993), and Dynamic Watershed Simulation Model (DWSM) (Borah et al., 2002). These models use climatic, physiographic (e.g., elevation, land use, soil), and manure or inorganic fertilizer application data to model the intensity and phosphorus concentration of runoff and phosphorus transport using a series of physics-based governing equations (Yang et al., 2016; Liu et al., 2020).

The model parameters, which control the simulated regional phosphorus contributions together with the input data, are calibrated against flow and water-quality measurements. Calibrated models can quantify the contribution of a certain source type, such as manure, by switching off its input and calculating the changes in the simulated phosphorus load. While using hydrologic models to simulate the flow dynamics is efficient, which we incorporate into our model framework, these models become significantly more computationally expensive and involve larger number of tuned parameters when involving multiple nutrient sources and transport processes. They are also cumbersome to deploy at the basin scale and require continuous updating as new water-quality measurements become available. The heavy reliance on a great variety of input data also makes these hydrologic models unsuitable for areas with limited data availability.

Instead, existing government assessments utilize simpler, data-driven approaches. It is valuable to distinguish between output- and input-based approaches, which differ primarily in the data they rely on for source attribution and can lead to substantially different results. Output-based approaches rely on existing water-quality measurements from waterways (e.g., Ohio EPA, 2016). The phosphorus contributions of a region bounded by the corresponding water-quality monitors can be derived using the measurements. However, in a given watershed, water-quality monitors with continuous observations tend to be sparse and non-uniformly distributed, leading to large and inconsistently sized attribution regions. Consequently, output-based approaches are inevitably limited in their ability to identify spatial variability in pollution.

Input-based approaches (e.g., ELPC, 2014; EWG, 2021) estimate excess phosphorus using a nutrient mass balance formula that subtracts crop uptake from phosphorus inputs. The phosphorus inputs and uptake by crops are constrained by data on manure production, fertilizer application, land use, and crop yield. Excess phosphorus estimates are generally available at annual intervals and are used as a proxy for a region's phosphorus contribution to the waterways (ELPC, 2014; EWG, 2021). As both the application of fertilizer or manure and the transport of excess nutrients during phases of high precipitation are seasonal, there can be significant deviations between the annual mean contributions and peaks within shorter time periods. In addition, input-based approaches implicitly assume that manure is applied to provide nutrients for cropland. In reality, however, there may exist illegal direct disposal of manure to waterway or spill of manure ponds, which should be prioritized in environmental enforcement but can be overlooked by input-based approaches.

To avoid specific assumptions about the level of fertilizer and manure application, we adopt a probabilistic model framework using ABC. Like other Bayesian approaches, ABC requires the inputs to have probability distributions (priors) from which inputs are sampled to identify distributions of outputs (posteriors) consistent with observations (Beaumont et al., 2002). The priors are constructed following the input-based approaches of excess phosphorus using manure production data, fertilizer application data, crop phosphorus uptake information, and flow dynamics in each subwatershed. Then we update the priors with water-quality measurements via ABC. The synergy of these data sources enables us to achieve improved spatiotemporal resolutions, accuracy, and efficiency over existing approaches. In this study, we develop the model framework for part of Western Lake Erie as a proof of concept, but our proposed method of combining data, hydrological modeling, and ABC can easily be implemented in other regions.

We focus on Lake Erie, as it has been experiencing recurring eutrophication and harmful algal blooms throughout recent decades, threatening the water supply for more than 12 million people in the U.S. and Canada (Michalak et al., 2013). The 1978 Great Lakes Water Quality Agreement (International Joint Commission, 1978) and subsequent regulation of point sources in the past led to a decline in algal blooms in Lake Erie by the 1980s (Kane et al., 2014). However, eutrophication and subsequent toxic algal blooms returned in the 1990s due to increased agricultural phosphorus runoff (Scavia et al., 2014), leading to low oxygen availability for fish and secretion of toxic material (Bridgeman et al., 2012). To address this crisis, the U.S. and Canadian governments agreed to reduce nutrient release by 40% by 2025 (Botts & Muldoon, 2005; Mohamed et al., 2019). Among several watersheds contributing nutrients to western Lake Erie, the Maumee River Basin has been identified as the largest contributor (Scavia et al., 2014; Bingham et al., 2015).

The Maumee River Basin (referred to as Maumee hereafter for simplicity) is the largest basin (16460 km^2) draining to Lake Erie, covering parts of Ohio, Michigan, and Indiana. The Lower Maumee River near the city of Toledo is its outlet. The Maumee River has five major tributaries: the St. Joseph, St. Marys, Auglaize, Blanchard, and Tiffin Rivers (Figure 1). Maumee has a hot-summer and humid continental climate, with most rainfall in March through July and snowfall in December through March. More than two-thirds of Maumee is cropland dominated by corn and soybean with sparsely distributed urban areas, pasture land, and forests. The soil in the region, composed primarily of silt, clay, and fine sand, has poor drainage capacity with high runoff potential (Myers et al., 2000). However, widespread tile drainage increases the drainage capacity of much of the cropland.

Maumee has seen a proliferation of permitted and unpermitted CAFOs over the last 30 years: Only 5% of the current (2019) CAFOs were constructed prior to 1990, with 43%, 35% and 17% built during each of the subsequent three decades (EWG, 2019). Maumee mainly contains swine, dairy, poultry, and cattle CAFOs, which generate vast quantities of liquid and solid manure. Manure and inorganic fertilizer applied to agricultural lands are major sources of phosphorus in the rivers of Maumee, which is the limiting nutrient for the formation of algal blooms in Western Lake Erie.

At the moment, several attribution attempts adopt purely data-driven approaches without accounting for pollutant transport. For example, one leading report estimates excess phosphorus in Maumee (ELPC, 2014) by comparing phosphorus input and uptake by crops (Stackpoole et al., 2019). We improve on such approaches by integrating flow dynamics that enable us to account for seasonal and spatial variability at the subwatershed scale. This framework, integrating data with nutrient transport, is poised to evolve and improve as more data and detailed physics for nutrient transport become available. While continued development is needed, the model is useful for permitting and targeted enforcement aimed at ensuring better compliance with existing regulations for surface water quality.

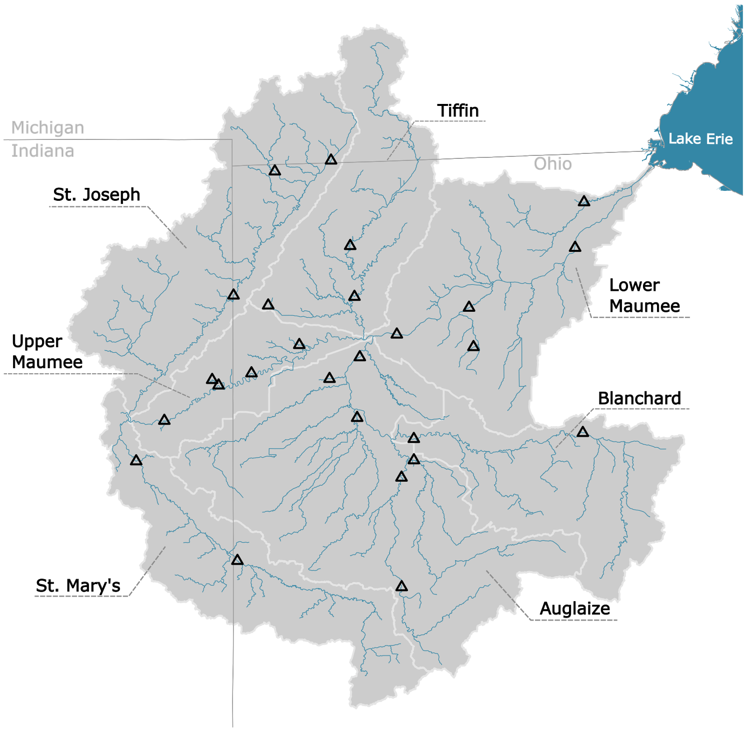

Figure 1. The Maumee River Basin. Seven HUC-8 watersheds are shown with white boundary lines. The watershed outlet is at Lake Erie on the eastern side. The basin is part of three states: Ohio, Michigan and Indiana. The USGS water-quality measurement locations are shown with black triangles.

2 Methodology

We use network modeling, hydrologic modeling, and Bayesian techniques to quantify the nutrient mass from different subwatersheds at high temporal resolution. In this study, we focus on the two forms of phosphorus, the organic or particulate form called unreactive phosphorus (UP) and the soluble inorganic form called soluble reactive phosphorus (SRP). We then estimate the relative contributions of manure, fertilizer, and soil to total SRP and UP. Figure 2 illustrates the architecture of our model. Table 1 defines key variables and parameters.

2.1 Data

Table 2 shows all data used in this study. We draw upon three broad categories of data—hydrologic, physiographic, and agricultural management data. Hydrologic data includes river discharge, stream network, and climate data. Physiographic data includes land use and soil type maps. Agricultural-management-related data includes fertilizer application rates; information about CAFO animal type, size, and count (used for manure estimation), and crop yield data. All these data were directly or indirectly used to general the network model or its inputs. We choose to prototype our model for the year 2019, one of the years for which the phosphorus data from the water quality monitor stations is the most complete.

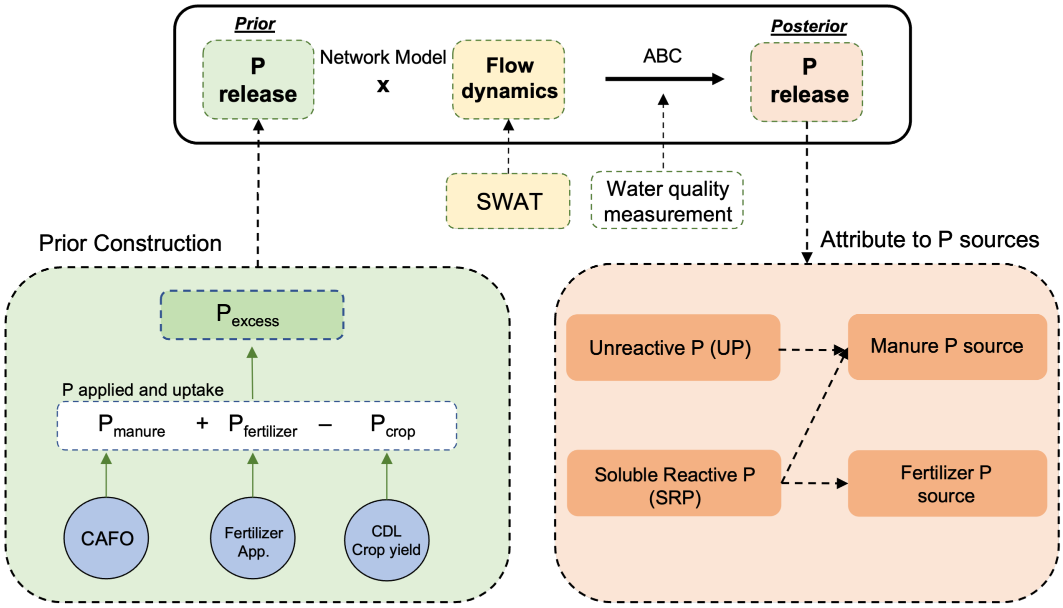

Figure 2. Model architecture. The central component is the model framework comprising the network model, which takes prior distributions and flow dynamics as inputs for the forward-modeling of nutrient transport, and ABC, which generates posterior distributions. Prior distributions are constructed using data on CAFOs, fertilizer application, and crop type, area, and yield.

2.2 Network Model

In discrete mathematics, a network or graph is a structure consisting of a set of points called nodes where each pair of nodes that share a given relationship is connected by a line, called an edge. These edges can be directed (e.g., river flowing from an upstream to a downstream node) or undirected (e.g., road connecting two cities). These simple building blocks can be used to construct network models representing interconnected systems in the extensive fields of social, natural, and engineering sciences (Khuller & Raghavachari, 1996; Chinowsky et al., 2008; Pokorádi, 2018). For a inland river system unaffected by tidal force, we choose to abstractly represent it as a directed acyclic network model, where water flows along directed edges and connects at junction nodes, but cannot flow back to a point upstream.

In this study, we represent the surface water system of Maumee using a network model where the subwatersheds are represented by source nodes, water quality monitors by monitor nodes, river confluences by junction nodes, and rivers by edges. Figure 3 shows a schematic of the network model. Each source node receives incoming nutrient load and adds its nutrient contribution. We assume the conservation of mass, thus the nutrient contributions of source nodes are non-negative. The monitor nodes provide locations for comparing simulated nutrient load with water quality measurements without modification. The junction nodes combine incoming nutrient load from upstream branches.

To construct the network model, we first simplify the stream network (USEPA & USGS, 2012) and divide branch bounded by confluences or monitors into segments, such that the area of land draining to the outlet of each segment is approximately at the HUC-12 scale (see supplementary information for details). The corresponding drainage area of each segment outlet forms a subwatershed in our model. Then we insert monitor nodes and junction nodes into the simplified stream network at the locations of water quality monitor station and river confluences, respectively. We place a source node at the out-

Table 1. Definitions and units of key variables and parameters.

Name	Definition	Unit
Network model		
S	Set of source nodes	
Q	Set of monitor nodes	
$D_q()$	Forward modeling function mapping sources S to monitor node q	
D_q^o	Observed nutrient mass at monitor node q	
Approximate Bayesian Computation (ABC)		
p_s	Prior distribution of nutrient concentration	
$W_{s,t}$	Water yield from source node s at time t	m^3
θ	An individual sample: a $\|S\| \times T$ matrix where each entry $\theta_{s,t}$ contains the mass at source node s at time t	g
t	Time index	days
T	Total simulation time period	days
N	Number of samples drawn in ABC	
n	Number of samples accepted in ABC	
d_q	Relative ℓ_1 distance between modeled and observed mass at monitor q	
w	Length of simulation window	days
Prior distribution		
m	Excess nutrient mass	g
C	Set of all CAFOs	
Relative contributions of manure, fertilizer, and baseline soil		
U	Mass of UP contribution of a subwatershed	g
R	Mass of SRP contribution of a subwatershed	g

let of each subwatershed, wherein the nutrient contribution of each source node is attributable to the corresponding subwatershed. As a result of this division, part of the subwatershed outlets and the locations of their corresponding source nodes overlap with monitor and junction nodes. The node relationships and resultant network model structure are illustrated in Figure 3. The length of the edge connecting each node is defined to be the length of the adjoining channel. We note that the network model facilitates a useful abstraction: It represents each subwatershed, which is a nonpoint source, as a single node in the network.

The network model domain considered in this study precludes the downstream lower Maumee river watershed represented as the empty portion in Figure 3, where algae consume significant quantities of nutrients for growth and form most algal blooms at Maumee (EWG, 2022). In 2019, the measured phosphorus load at the outlet of the lower Maumee River watershed was lower than its incoming nutrient load. To ensure conservation of mass remains a valid assumption, we choose to exclude the lower Maumee River watershed from our model domain. Therefore, the network outlet is the monitor node just upstream of the lower Maumee River (Figure 3).

The complete network model of Maumee comprises 489 edges and 490 nodes with 328 source nodes, 142 junction nodes and 20 monitor nodes (see Figure 3). We let S de-

Table 2. Types and sources of data used in the current study. Sources listed in the table include the National Center for Water Quality Research (NCWQR), National Hydrography Dataset (NHD), United States Geological Survey (USGS), Environmental Working Group (EWG), National Agricultural Statistics Service (NASS), Soil Survey Geographic database (SSURGO), Oregon State University (OSU), and Oak Ridge National Laboratory (ORNL).

Type	Source	Spatial	Temporal	Reference
Network model setup				
Water quality	NCWQR, USGS	26 stations	Daily	(NCWQR, 2022)
River discharge	USGS	58 stations	Daily	(USGS, 2016)
Stream network	NHDPlusV2	HUC-12	Present	(USEPA & USGS, 2012)
Inputs to prior formulation				
CAFO	EWG	Point	1988-Present	(EWG, 2019)
Fertilizer rate	USGS	County level	2002-2017	(Falcone, 2021)
Land use and crop	USDA-NASS	30-m	2002-2021	(Boryan et al., 2011)
Crop yield	USDA-NASS	State level	2006-2021	(USDA-NASS, 2021)
Climate data				
DAYMET climate	ORNL	1km	1980-Present	(Thornton et al., 2016)

note the set of source nodes and Q denote the set of monitor nodes in the network. For the network model of Maumee, $|S| = 328$ and $|Q| = 20$.

We route nutrients through the network via advection. Here, we use the edge lengths l (m) and hourly channel velocity time series $v(t)$ (m/s) along each edge, which are interpolated from daily SWAT velocity estimates. We compute the time l/v for nutrients departing each upstream node at a given hour to arrive at each downstream node, where we assume that nutrients move at the same velocity as the water in the channel. With these travel times, we construct the forward-modeling function $D_q()$, which maps the input nutrient mass departing each source node $s \in S$ to compute the total mass arriving at each monitor node $q \in Q$ over each time step $t \in T$. We compute the observed mass at the monitor node by multiplying the observed daily concentration (g/m^3) and daily discharge (m^3/s) and scaling by 24×3600 to obtain the total daily observed nutrient mass. We denote the time series of daily observed nutrient mass at monitor node q as D_q^o.

2.3 Approximate Bayesian Computation

Approximate Bayesian Computation (ABC) is a rejection-based computational method for calculating posterior distributions of unknown model parameters (Beaumont et al., 2002; Csilléry et al., 2010; Sunnåker et al., 2013). In our implementation of ABC, samples of source nutrient contributions are accepted/rejected based on the difference between simulated and observed nutrient loads. ABC is mathematically simple but robust, without relying upon more complex likelihood functions like fully Bayesian methods (Sunnåker et al., 2013). Using ABC, we can extensively test possible values in the prior distributions of inputs without falling into local minima. ABC is particularly suitable for our study because (1) the rapid forward modeling of nutrient transport through the network makes possible the large number of samples and simulations required due to the large number of sources; (2) the method is robust for both uninformative, poorly constrained

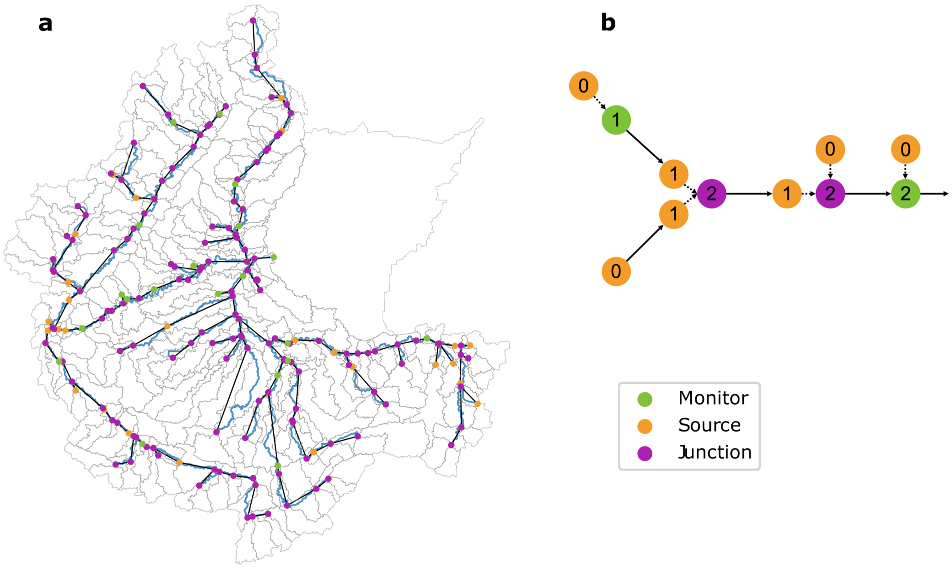

Figure 3. Network model representation of the stream network at Maumee, with monitor, source, and junction nodes shown with green, yellow and purple points respectively, and edges shown with black arrows. **(a)**: Overview of the entire network model, subwatersheds, and major channels (blue lines). For readability, the source nodes overlapping with junction and monitor nodes are not shown. The empty portion on the right depicts the lower Maumee river watershed. **(b)**: Illustration of node relationships present in the network model. The number on each node represents the number of its incoming edges. Arrows represent edges. Solid arrows represent channels, while dashed arrows represent node connection with zero physical length. All nodes have 1 outgoing edge except that the basin-outlet monitor node has none. Upstream-most source nodes have 0 incoming edge, while the others have 1. All monitor and junction nodes have 1 and $\geq$2 incoming zero-length edges from upstream source nodes, respectively (hidden in Figure 3a). Upstream-most monitor nodes only receive nutrient contribution from its associated source nodes and have 1 incoming edge, while the others also receive upstream nutrient load and have 2 incoming edges.

(e.g., uniform) and informative, well constrained (e.g., data-driven) priors; and (3) the generated posterior distribution naturally enables uncertainty quantification.

We use ABC to sample nutrients contributed by each source node. Note that ABC is performed independently for each nutrient, so we describe the process for a single nutrient. For each source node $s \in S$, we define a distinct prior distribution p_s over the nutrient concentration. The derivation for p_s is described in detail in Section 2.5. We generate an input mass sample at source s and daily time step t by sampling a concentration from p_s, and multiplying by the daily water yield $W_{s,t}$. The water yield is an output from the SWAT model, and is representative of the total outflow from a subwatershed.

However, as nutrients from different source nodes are aggregated across time and space in the simulation, an independent ABC sample does not merely consist of a sampled mass at a given day and source. Rather, a sample $\theta \in \mathbb{R}^{|S| \times T}$ is a matrix, where a given entry $\theta_{s,t}$ is the mass sampled for a particular source s and day t, and T is the number of daily time steps in the simulation. We generate N samples from the prior distributions and run the forward modeling process D_q with each sample θ, generating N sets of outputs for each monitor $q \in Q$. Each output of size $\mathbb{R}^T$ represents time series of the simulated nutrient load at a given monitor. At each monitor node q, we compare the sample output, $D_q(\theta) \in \mathbb{R}^T$, and observations, $D_q^o \in \mathbb{R}^T$, by computing the relative ℓ_1 distance d_q:

$$d_q = \sum_{t=1}^{T} \frac{|D_{q,t}(\theta) - D_{q,t}^o|}{P_{99}(D_q^o)}, \tag{1}$$

where $P_{99}(D_q^o)$ denotes 99^{th} percentile of the observed daily time series, which we divide by to normalize the distances at each monitor node, thus weighting each monitor node equally. We use the 99^{th} percentile to trim outliers. We note that when a observed value $D_{q,t}^o$ is missing, the given term is ignored in the summation. We accept the n samples resulting in the smallest average distance over all monitors. The accepted samples generate the posterior distributions of the nutrient input of each source node at each daily time step.

To increase computational efficiency and decrease the size of each ABC sample θ, we divide the full simulation period $T = 365$ into smaller portions. We fix a target simulation window of w time steps over the observed monitors, and determine the source days such that nutrients departing these sources would arrive at a downstream monitor within the observed simulation window. Thus, we run T/w independent simulations, retaining only accepted samples for relevant source days. Note that this means that each source day posteriors are comprised of accepted samples from multiple simulation windows. In this study, we choose $N = 10^5$, $n = 10$, and $w = 1$. Higher N slows down the model without significantly increasing the model performance.

2.4 Hydrologic Model

The network model requires subwatershed-scale flow dynamics as an input to calculate nutrient load. Here we used the Soil and Water Assessment Tool (SWAT), a physically based, semi-distributed hydrologic modeling software (Arnold et al., 1998) to simulate the flow dynamics. The SWAT model uses climate forcing data and physiographic data (e.g., soil and land use), and it solves the water balance equation to estimate hydrologic components like surface and subsurface flow, which is then used to estimate streamflow. Note that our model framework only requires running SWAT once, where we calibrate and validate the model for the years 2015-2020 at Maumee and simulate the flow dynamics. We then use the pre-computed subwatershed-level water yield and channel velocity as inputs to the network model. Details about the SWAT model are included in the supporting information.

2.5 Prior estimation

The network model uses an informative prior in ABC for source nodes, where each node s represents a subwatershed. For each subwatershed s, we select a beta prime prior distribution centered at its estimated excess phosphorus. In the following sections we describe the methods to estimate excess phosphorus and the parameterization of the prior distribution.

2.5.1 Excess phosphorus estimation

We estimate excess phosphorus at the subwatershed scale by solving phosphorus mass balance over land. The source term in the phosphorus mass balance formula are the phosphorus input from manure and fertilizer application, whereas the sink term is the uptake of phosphorus by crops. We first estimate the annual excess phosphorus mass in subwatersheds and then divide it by the annual water yield from the SWAT model to calculate the concentration. We construct priors separately for UP and SRP. We assume that manure contributes to both UP and SRP, inorganic fertilizer contributes to only SRP, and plants consume only SRP. Therefore, we estimate excess UP of subwatershed s, U_s, based on the manure application to the agricultural land,

$$U_s = U_s^m, \tag{2}$$

where U_m^s is the total mass of UP from applied manure in subwatershed s. On the other hand, we estimate excess SRP of subwatershed s, R_s, based on inorganic fertilizer application, manure application and plant uptake,

$$R_s = R_s^m + R_s^f - R_s^k, \tag{3}$$

where total mass of SRP in subwatershed s are input as applied manure, R_s^m, and applied fertilizer, R_s^f, and output as crop uptake R_s^k.

Specifically, we estimate the manure phosphorus (UP or SRP) from each CAFO by the product of animal population, manure produced per animal, and manure phosphorus content. We follow EWG (2019) and EWG (2021) and set different manure production rates and phosphorus contents for each major CAFO animal type at Maumee: dairy, cattle, swine, and poultry. Then, assuming the manure is evenly applied to cultivated cropland and pasture within a 5-mile buffer around each CAFO, we calculate the manure phosphorus of a subwatershed by aggregating the intersecting proportions of all CAFO buffers with this subwatershed. The assumed 5-mile application range is supported by previous studies showing that most manure is applied within short distance around CAFOs (Long et al., 2018; Kast et al., 2019). Without existing analysis on different application range of different manure types, we utilize a constant radius for all CAFOs for simplicity. We calculate the cropland area using the 30-m Cropland Data Layer from the United States Department of Agriculture (Boryan et al., 2011). Mathematically,

$$P_s^m = \sum_{c \in C} a_s^c \gamma_P^c, \tag{4}$$

where P denotes either UP or SRP, C is the set of all CAFOs, a_s^c is the area of subwatershed s where the cultivated cropland and pasture intersect the manure application buffer of a CAFO c, and γ_P^c is the spatial density of UP or SRP for $c \in C$, defined as:

$$\gamma_P^c = \frac{m^c \phi_P^c}{\sum_{s \in S} a_s^c + a_e^c}, \tag{5}$$

where m^c is the manure mass from c, ϕ_P^c is the weight percentage of UP or SRP in the manure type of CAFO c, and a_e^c is the area of cultivated cropland and pasture outside Maumee that intersects the manure application buffer of CAFO c. We calculate ϕ_P^c following EWG (2021) based on the manure composition data by Barnett (1994) and EWG (2019).

We estimate SRP from inorganic fertilizer for subwatershed s by multiplying the application rate by cultivated cropland area, assuming inorganic fertilizer provides only SRP (Kleinman et al., 2002; Culman et al., 2020). We use county-level inorganic fertilizer application rates over the conterminous U.S. provided by USGS (Falcone, 2021). Mathematically,

$$R_s^f = a_s \gamma_{s,R}, \tag{6}$$

where a_s is the cultivated cropland area in s, and $\gamma_{s,R}$ is the spatial density of fertilizer SRP application in s.

We estimate subwatershed-scale crop SRP uptake based on the yields (USDA-NASS, 2021), areas (Boryan et al., 2011), and phosphorus uptake rates (Watters, 2021) of different crop types. Mathematically, the SRP uptake in subwatershed s is

$$R_s^k = \sum_{i \in I} a_s^i y_s^i k^i, \tag{7}$$

where I is the set of crop types, and a_s^i and y_s^i are the area and yield in s of crop type i respectively, and k^i is uptake rate of crop type i. In this study, I contains corn, soybean, wheat, alfalfa, and other hay.

2.5.2 Prior Distribution

We assign each subwatershed source node s with data-driven prior distributions of nutrient concentration. Specifically, we sample nutrient concentrations and multiply them with subwatershed-scale water yield time series to acquire the nutrient mass inputs time series, which are then transported in the network. We use the beta prime distribution as the prior distribution p_s of the nutrient concentrations for source s. The probability density function is defined as:

$$p_s(x) = \frac{x^{\alpha-1}(1+x)^{-\alpha-\beta_s}}{B(\alpha, \beta_s)}, \tag{8}$$

where $x > 0$ is the nutrient concentration, B is the beta function, and α and β_s are the two parameters of the distribution, where α is a chosen hyperparameter and β_s varies by subwatershed.

We center the prior distribution p_s for each nutrient at the estimated excess phosphorus concentration for subwatershed s derived in Section 2.5.1. Then we solve for the parameter β_s using the expectation of nutrient concentration over the subwatershed prior $\mathbb{E}(x) = \frac{\alpha}{\beta_s - 1}$ (if $\beta > 1$), yielding

$$\beta_s = \frac{\alpha \sum_t^T W_s^t}{U_s} + 1 \tag{9}$$

for β_s for UP. This calculation is defined identically for SRP. We fix $\alpha = 0.8$ for UP and $\alpha = 0.5$ for SRP, where these parameters are chosen to encourage a large mass near 0 (particularly for the smaller valued SRP), while still allowing for a reasonable probability of sampling larger values.

2.6 Relative contributions of manure, fertilizer, and baseline soil

To determine the relationship of UP and SRP to manure, inorganic fertilizer, and baseline soil phosphorus, we develop a procedure illustrated in this section, leveraging previous experimental results (Sharpley, 1997; Kleinman et al., 2002).

About half of phosphorus in both liquid and solid manure is UP in organic or particulate forms (Fordham & Schwertmann, 1977; Barnett, 1994; Kleinman et al., 2002; J. C. Hansen et al., 2004). In contrast, the dominant form of phosphorus in inorganic fertilizer, such as monoammonium and diammonium phosphate (Culman et al., 2020), is phosphate (e.g., Kleinman et al., 2002)—i.e., SRP. According to the runoff experiments by Kleinman et al. (2002) and Bertol et al. (2010), UP concentrations in runoff with and without application of inorganic fertilizer are similar, while UP concentrations in runoff with manure application is significantly higher than the control and fertilizer groups by a factor of 2. Therefore, in this study where we consider the short-term (weeks to months) effect of fertilizer and manure application on nutrient loss, we assume that no phosphorus from inorganic fertilizer becomes UP and thus only manure application increases UP concentration in runoff (i.e. $U^f = 0$), yielding the following relationship:

$$U = U^m + U^l, \tag{10}$$

where U^m and U^l denote the contributions of UP mass by manure and soil respectively. The contribution of soil is a function of the baseline soil phosphorus level, which depends on soil type, the long-term application intensity of manure and fertilizer, and the rate of phosphorus removal via crop uptake or runoff.

To calculate U_m from the UP obtained from the network model, we first estimate UP from soil, U_l. Kleinman et al. (2002) conducted controlled experiments with high-P and low-P soils and found that UP concentration in runoff is sensitive to soil phosphorus level. Although we lack data for constructing quantitative relationship between soil phosphorus level and the concentration of UP in runoff, the measured Mehlich-3 P of the soil samples used in Sharpley (1997) is similar to the county level median Mehlich-3 P at Maumee in 2015 (Dayton et al., 2020). For example, the median Mehlich-3 P levels of Auglaize County in Ohio in 2015 and the soils used in Sharpley (1997) are 33 mg/kg and 25 mg/kg, respectively. However, according to Dayton et al. (2020), the Mehlich-3 P of samples within counties are highly varied. We acknowledge our estimation is first-order, with the uncertainty associated with the spatially coarse and temporally sparse soil phosphorus data and the lack of direct measurements for runoff phosphorus concentration at Maumee. For each subwatershed s at time step t,

$$U^l_{s,t} = \min(W_{s,t}[U]^l, U_{s,t}), \tag{11}$$

where $[U]^l$ is the mean UP concentration reported in the control experiments of Sharpley (1997), and $U_{s,t}$ is the UP mass estimated by the network model. We then calculate U^m using Eq. (10) and U^l acquired in the first step.

After calculating the UP contribution of manure for each source and time step, $U^m_{s,t}$, we calculate the SRP contributions by soil and manure. We first calculate the SRP contribution of soil, $R^l_{s,t}$, in the same way as UP using Eq. (11). Then we calculate the SRP contribution of manure, $R^m_{s,t}$, based on manure compositions. The forms of phosphorus in manure vary with manure forms and animal types. We use the mass ratio $SRP/UP = \lambda = 0.98$ based on the mean value of the data reported in Barnett (1994) to calculate the SRP contribution by manure

$$R^m_{s,t} = \min(\lambda U^m_{s,t}, R_{s,t} - R^l_{s,t}). \tag{12}$$

Therefore, the SRP contribution by inorganic fertilizer is

$$R^f_{s,t} = R_{s,t} - R^m_{s,t} - R^l_{s,t}. \tag{13}$$

3 Results

3.1 Improving spatial and temporal inferences in phosphorus release

Existing methods that mostly rely on data for quantifying the phosphorus released from different regions in a given watershed are spatially and temporally coarse (e.g., ELPC, 2014; EWG, 2021). As discussed in the introduction section, input-oriented methods like ELPC (2014) and EWG (2021) provide estimates at a relatively fine spatial scale but only on an annual basis. Output-oriented methods (e.g., Ohio EPA, 2016) relying primarily on water-quality measurements allow for high temporal variability but at a relatively coarse spatial scale. Recognizing the complementary nature of these two existing approaches, our model combines both data sources to improve our ability to draw spatial and temporal inferences.

Figure 4 compares the spatial variability in estimated unreactive phosphorus (UP) and soluble reactive phosphorus (SRP) density over 2019 using three different approaches. We focus on the year 2019 as a proof of concept, because it has more data available than earlier years and is not yet confounded by the onset of the COVID-19 pandemic. The left column (Figures 4a and 4d) mimics an output-oriented approach as used by Ohio EPA (2016) with our estimation using only spatially sparse water quality time-series. The

Figure 4. Spatial distribution of UP and SRP release. **(a,d)** Coarse scale output-based attribution using only water quality observations with watersheds delimited by monitors (black circles). **(b,e)** Fine scale attribution leveraging CAFO (white points), fertilizer, and crop data to compute annual excess phosphorus. **(c,f)** Fine subwatershed scale attribution using the network model and ABC, which integrates the two approaches.

middle column (Figures 4b and 4e) represents an input-oriented approach as employed by ELPC (2014) using spatially granular estimates of excess phosphorus on the annual scale. The right column (Figures 4c and 4f) shows the network model output that incorporates both water quality time-series and excess phosphorus data used in the output- and input-based estimates shown in the first two columns of Figure 4, respectively.

The differing spatial resolution of output- and input-based approaches is evident from the degree of variability in estimated phosphorus release in Figures 4a and 4d as compared to Figures 4b and 4e. In Figures 4a and 4d the watersheds are defined based on the location of water-quality monitors, yielding 23 regions bounded by the 23 monitor locations (USGS, 2016; NCWQR, 2022) depicted as black circles in Figure 4a. Due to the long distance between monitor nodes, most of the output-based watersheds are large. In contrast, the input-based watersheds in Figures 4b and 4e are subwatersheds. In size, these subwatersheds are comparable to USGS HUC12 scale watersheds. Our model (Figures 4c and 4f) maintains this subwatershed-scale resolution by using highly variable excess phosphorus estimates as a prior, but additionally leverages existing measurements of water quality over time to update the prior, primarily in regions where estimated excess phosphorus mismatches the observed phosphorus load.

In Figures 4a and 4d, we first estimate the annual phosphorus load from daily time series at the inlet and outlet monitor nodes of each watershed. Then we divide the difference by the area of the watershed to estimate annual phosphorus release density. A striking feature of the resulting output-based UP estimates (Figure 4a), is a homogeneously high UP release density in the lower half of the domain, primarily in the watersheds of

St. Marys and Auglaize (for the exact boundaries of these watersheds, see Figure 1). However, the highest UP release density (>280 kg/year/km^2) is attributed to the two smallest watersheds in upper Maumee with areas less than 50km^2. The estimated SRP release density in these two watersheds is also about twice as high as in the surrounding areas, but attains its maximum value in upper St. Marys (see panel d).

Figures 4b and 4e show input-based estimates of excess phosphorus release density, where the finer scale attribution is facilitated by the high spatial resolution of the input land use and CAFO data (see Table 2). We estimate excess phosphorus mass using cropland, CAFOs, and county-level fertilizer application data by subtracting crop uptake from manure and inorganic fertilizer inputs. Then we divide the excess phosphorus mass by the area of subwatershed to calculate the subwatershed-scale phosphorus density estimates. Higher excess UP indicates higher availability of organic and particulate which are primarily sourced from CAFO manure, while higher SRP is more indicative of higher inorganic fertilizer application. Although fertilizer directly contributes to SRP, about half of manure P is also SRP (Barnett, 1994). Therefore, high CAFO manure production and high inorganic fertilizer application can both lead to high SRP contribution.

The input-based approach entails great spatial variability in excess phosphorus estimates, even for neighboring watersheds. Figure 4b shows high excess UP (> 300 kg/year/km^2) availability in St. Marys, Upper Maumee, upper St. Joseph and Tiffin, and pockets of Auglaize—all areas with particularly high CAFO density as evident in Figure 4b where CAFOs are represented as white dots. In contrast, Figure 4e suggests that several large regions including southern St. Joseph's and western Blanchard release very low SRP, while very high excess SRP is found throughout Tiffin, along the southwestern border of St. Marys, upper St. Joseph, and northern Auglaize. The spatial contrast in estimated phosphorus levels between neighboring subwatersheds is higher for SRP than for UP and tends to occur between neighboring subwatersheds with differences as high as 1700 kg/year/km^2. The spatial contrast also coincides with vertical county boundaries at some regions, such as upper St. Joseph and Auglaize, as a result of using the county-level fertilizer application rates (Falcone, 2021).

Finally, Figures 4c and 4f show the fine subwatershed-scale attribution using our model, in which we draw phosphorus samples from a prior distribution of excess phosphorus and route these through the stream network using the simulated flow information from the SWAT model, but only retain samples that match the observed water quality measurements. Our model maintains a similar spatial resolution as the input-based approach (Figure 4b) and pinpoints possible regions of peak contribution more specifically than the output-based approach. The model estimates are broadly consistent with the output- and input-based approaches in the sense that portions of the upper Maumee and St. Marys watersheds are expected to contribute the highest UP levels (Figure 4c), but rather different in the details. In particular, our model reduces the spatial contrasts in the UP and SRP contributions between neighboring subwatersheds, especially in the vicinity of high-contribution subwatersheds.

The differences between our model estimates (Figures 4c and 4f) and the other two approaches begs the question why the estimates differ. Comparing our model to the output-based approach first, one issue is that the monitor-delimited watersheds in Figures 4a and 4d differ by more than two orders of magnitude in size, spanning areas from 10km^2 to 1560km^2. The two watersheds attributed with the highest UP release density are among the smallest watersheds (<50km^2), suggesting that the heterogeneous sizes of the watersheds may bias estimates: There are potentially other small high-density regions within larger low-density regions, but when aggregated over a large area, the contributions of small regions are smoothed out.

The highly heterogeneous attribution suggested by the input-based approach supports the previous argument that the output-based approach is smoothing out extreme values. However, some of these high values and discontinuities may be the result of the assumptions required to convert input data at various spatial resolutions to the subwatershed scale. While CAFO locations are points, cropland data is available in a 30-m resolution, and fertilizer application is estimated at the county level. Potential evidence of this issue is that the highest UP value of over 2000 kg/year/km^2 in Upper Maumee occurs at the intersection of overlapping manure application areas, each with an assumed average 5 mile radius. Similarly, sharply contrasting estimates sometimes correlate with county boundaries that are unlikely to cause drastically different farming practices such as the low-density western third of the basin, and the high-density eastern boundary of Blanchard in Figure 4e.

Our model attempts to strike a balance between these two prior approaches. It retains much of the spatial heterogeneity suggested by the prior. The additional information on phosphorus inputs allows the model to disaggregate the often large drainage area between two monitors into subwatersheds with high and low levels of expected phosphorus release. For example, the two monitor-delimited watersheds constituting St. Marys have an estimated UP density of 210 and 224 kg/year/km^2 in Figure 4a. Our model considering 74 different subwatersheds within St. Marys estimates UP density ranging from 32 to 324 kg/yr/km^2. Meanwhile, our model reduces inconsistencies between estimated phosphorus inputs and measured phosphorus in the streams, leading to a spatially smoother attribution. For example, large excess UP estimates in Upper Maumee and outlying excess SRP estimates on the western border of Blanchard decrease on average by over 50%.

The differential updating of expected phosphorus contributions flowing to different monitors suggests that our model is able to learn from the available water-quality data. In addition to providing a spatially more nuanced assessment of likely phosphorus release, our model resolves one fundamental disconnect between the two prior models, namely that the input-based model entails significantly higher levels of total phosphorus release than the output-based model. Overall, we find that the excess phosphorus estimated by the input-based model exceeds that of the output-based model by 29% and 156% for UP and SRP, respectively. By integrating the water quality observations into our model, this overestimation drops to 9% and 53%, respectively. A partial disconnection between excess phosphorus and phosphorus transport in streams is not necessarily unexpected, because processes such as manure storage, application approaches, phosphorus storage in the soil, soil erosion and land-use management alter how much phosphorus is applied and how it is redistributed after application.

To better understand the updates needed to improve the consistency with water-quality data, we compare the discrepancy between the prior (represented by Figures 4b and 4e) and the posterior (represented by Figures 4c and 4f) for all subwatersheds in Figure 5. We plot the mean of the posterior, representing the point estimate from our model, against the mean of the prior, representing the estimated excess phosphorus input, for each subwatershed at the annual scale. The points are colored by the immediate downstream monitor, and points falling below (above) the dotted black line represent watersheds in which the updated estimate is lower (higher). The majority of subwatersheds falls well below the no-update line, implying that the prior overestimates phosphorus contributions, particularly for SRP and subwatersheds with high contributions. The only area where the prior underestimated phosphorus release is Auglaize watershed for UP (Figure 5a). While the ABC decreases the prior UP and SRP estimates on average, the updates differ at different locations in the network, reflecting specific signals from the water-quality measurements.

Excess phosphorus estimates are generally limited to annual scale by data availability (e.g., ELPC, 2014), and thus any higher temporal dynamics in UP or SRP mass estimates are entirely reliant on flow patterns. From a practical point of view, it is un-

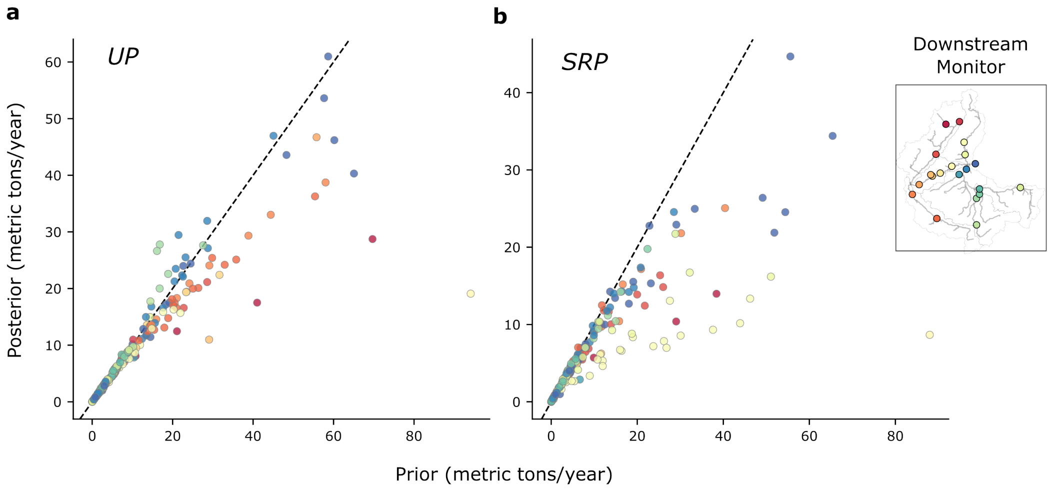

Figure 5. Annual excess phosphorus estimates (prior) vs. phosphorus attribution by the network model (posterior mean) of (a) UP and (b) SRP. Each point represents the annual release from a subwatershed, colored by the immediate downstream monitor as depicted by the map legend on the left. The dotted black line represents no updating, in that the the expectation of the prior and the posterior are equal.

realistic to assume the nutrient concentration remains same throughout the year, particularly in agricultural areas where seasonal farming patterns influence phosphorus release. Incorporating the water-quality measurement time series not only ensures that our model estimates are more consistent with the measurements, but also allows for fine-grained temporal attribution.

In Figure 6 we compare the daily time series of phosphorus load forward-modeled to monitor nodes as predicted by the network model posteriors against the input-based estimates. In the input-based estimates, the daily nutrient mass is proportional to the daily water yield, assuming constant nutrient concentrations throughout the year.

For concision, we only show estimates for two monitor nodes, with SRP shown for a low-flow monitor in Figure 6a and UP shown at a higher-flow monitor in Figure 6b. As we have already noted the overall upward bias in the input-based estimates in annual scale analysis, we choose to display time-series that exemplify the limitations of the brittle assumption of constant concentration: the inability to differentiate daily flow dynamics from pollution trends and the insufficiency to account for important seasonal cropping patterns. We note that the inferior fit by the prior shown in these two plots exemplifies the prior error. The average relative ℓ_1 error (see Eq. (1)) between the median of the prior and observed over all monitors is about 44%, and 26% of that of the phosphorus estimate for UP and SRP respectively.

Figure 6a demonstrates two key ways in which the network-model estimates outperform the input-based estimates in capturing SRP temporal dynamics. First, when SRP spikes at several points during the relatively lower flow winter time, the network-model estimates generally include the the peaks, although underestimating the actual contribution. The input-based estimates on the other hand, fail to capture these spikes, and significantly overestimates SRP load during January to June. Second, the recession pattern after the peak events are relatively slow in the input-based estimates, following the recession pattern of the flow. Such slow recession limb is not present in the observations or the network-model estimates.

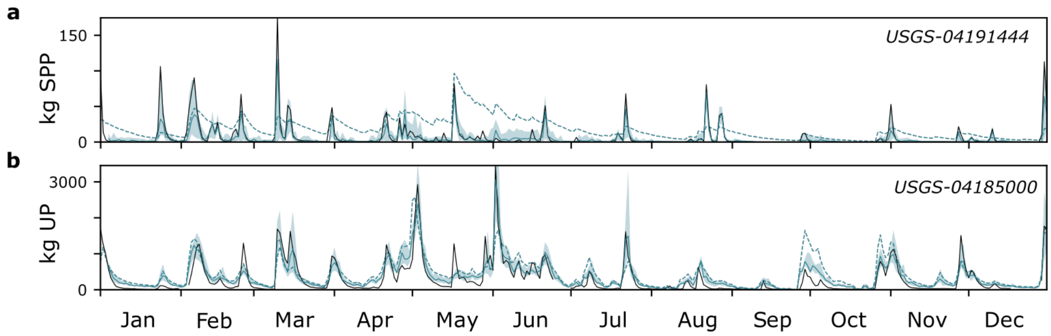

Figure 6. Time series of phosphorus mass for (a) SRP at a relatively low-flow monitor in Auglaize and (b) UP at a relatively high-flow monitor node in Tiffin. The network model 90% credible interval and median are depicted with a blue shaded region and solid blue line, respectively. input-based estimates are shown with a dotted blue line. The observed mass at monitor nodes is shown with a solid black line.

While the input-based and network-model estimates are much more closely aligned for UP at monitor USGS–04185000 as shown in Figure 6b, the network-model estimates still outperform the input-based estimates. Although the network model slightly overestimates UP load during lower flow periods (November-December), the 90% credibility interval of the posterior generally include the observation during high flow periods. In contrast, the input-based approach overestimates UP load during low flow periods in October and November specifically when there is a peak event, whereas it underestimates UP load during spring and summer peaks. These mismatches further indicate the missing temporal dynamics in the input-based estimates.

Although our model posterior is more consistent with the water quality observations than the excess phosphorus, it still overestimates the overall contributions. The high temporal variability in measured phosphorus loads shown in Figure 6 reveals the limitation of our model assumption and sampling approach that lead to the overestimation. As illustrated in section 2.3, we assume a constant daily input at each source node, which can affect the phosphorus loads of multiple days at downstream monitor nodes. When the water-quality measurements show sharp temporal variations, this assumption hinders the ability of our model to fully match the data. Moreover, the high dimensions of independent samples, each of which contains contributions of all subwatersheds, also add to the overestimation. At days with low phosphorus loads, among the computational viable number of samples, even the smallest sample can still be too high, especially with temporally constant prior distributions that significantly overestimate subwatershed contributions.

Overall, the above analysis underscores the significant limitations in the use of annual scale excess phosphorus to attribute phosphorus at high temporal frequency. The temporal analysis reveals that the issue with the excess phosphorus estimates is not merely overestimation that can be easily remedied by applying a scaling factor, but an overall lack of robustness in capturing temporal dynamics. This examination of the time-series posteriors also highlights the advantages of establishing a posterior distribution at each time step rather than a single time-series in capturing highly variable daily and seasonal trends.

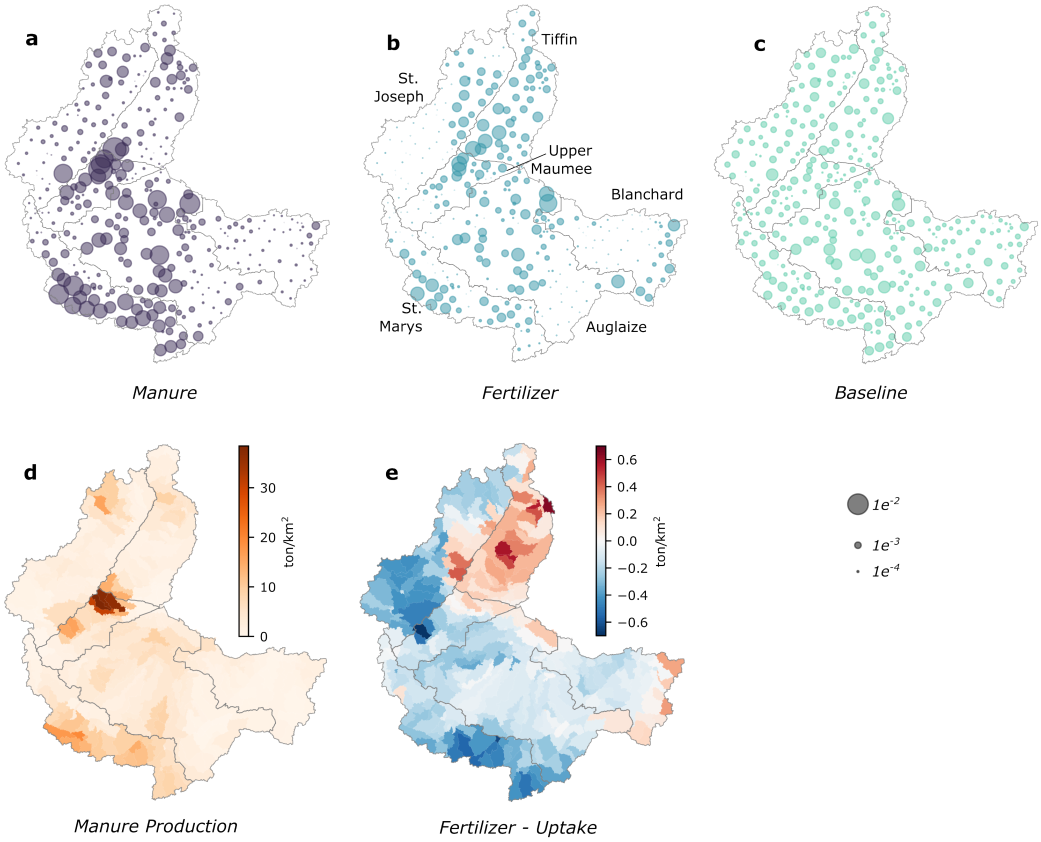

Figure 7. **(a-c)**: Spatial attribution of 2019 surface water phosphorus sourced from (a) manure (b) fertilizer and (c) baseline phosphorus, which includes comprises the soil phosphorus from fertilizer, manure, plant residual accumulated over the years. Each circle represent a fraction of the total annual total phosphorus in the surface water in the given area, where the size is proportional to the contribution. **(d)**: Subwatershed-scale plot of manure production (EWG, 2019). **(e)**: Subwatershed-scale plot of fertilizer phosphorus application (Falcone, 2021) subtracted by crop uptake (Boryan et al., 2011; USDA-NASS, 2021; Watters, 2021) in spatial density.

3.2 Manure contributes more phosphorus than fertilizer

Besides the fine spatial and temporal resolutions, identifying specific source types is a necessary component of phosphorus attribution intended for an actionable nutrient management plan. Most phosphorus entering the streams via rainfall or snow melt runoff is from manure and fertilizer widely applied throughout the basin. Part of this phosphorus is from newly applied manure and inorganic fertilizer on land surface before they are absorbed by soil, and the rest is from phosphorus accumulated in the soil from historical applications. We refer to the part of phosphorus from soil as "baseline phosphorus", which is present in runoff regardless of (Sharpley, 1997; Kleinman et al., 2002) and continuously replenished by (Nair et al., 1995) recent applications.

Figures 7a–c shows the spatial distribution of relative contributions of manure, fertilizer and baseline phosphorus as a fraction of total annual phosphorus release at Maumee. For each subwatershed, we first calculate the baseline UP and SRP by multiplying water yield with measured concentrations from runoff experiments with similar soil phosphorus level as Maumee (Sharpley, 1997; Dayton et al., 2020). Then we subtract the baseline phosphorus from the modeled UP and SRP illustrated in section 3.1 to compute the contribution of manure and fertilizer. Assuming fertilizer only contribute to SRP, we estimate manure UP as the remaining UP and calculate manure SRP using manure compositions. By subtracting the calculated manure SRP from the total remaining SRP, we then obtain fertilizer SRP (See section 2.6 for details). Note that the source type attribution is based on the modeled phosphorus entering the streams for 2019, and the estimates for manure and fertilizer shown in Figures 7a and 7b represent the contributions from application over 2019. The contribution of baseline phosphorus shown in Figure 7c, however, can include phosphorus accumulated from manure, fertilizer and plant residues from past years.

Figure 7a shows substantial spatial heterogeneity in the contributions of manure. Comparison between Figures 7a and 7d shows that the spatial pattern of phosphorus contribution by manure is highly consistent with that of manure production, indicating the high impact of the CAFOs to the total phosphorus release. However, the relative phosphorus contributions of subwatersheds, which is the attribution result of our network model, significantly differ from the relative magnitude of manure phosphorus production, suggesting that phosphorus contribution by manure depends on multiple factors, rather than just manure production.

Figure 7b shows that the contribution of fertilizer is also spatially heterogeneous but in a different way from manure. In some regions, such as St. Marys and upper Maumee, both manure and fertilizer show high contributions with locally similar spatial pattern (Figures 7a and 7b). According to Figures 7d and 7e, this pattern is likely a result of fertilizer application along with excessive manure application that results in loss of surplus phosphorus from both sources. In contrast, some other regions such as part of Tiffin and St. Joseph show high fertilizer but little manure contribution. These regions coincide with the regions with surplus phosphorus in Figure 7e and relatively low manure application rates in Figure 7d. Therefore, this high-fertilizer and low-manure spatial pattern may indicate excessive fertilizer application in regions without significant manure application.

Figure 7c shows the significant and relatively homogeneous baseline phosphorus contribution throughout Maumee. It indicates that the baseline phosphorus contribution, which is a result of long-term accumulation of phosphorus from different sources, is also an important contributor of total phosphorus at Maumee. The homogeneity of the inferred baseline phosphorus stem from the our assumption of constant baseline UP and SRP concentrations based on experimental data (Sharpley, 1997). In regions where the contributions of both manure and fertilizer are low, such as Blanchard, lower St. Joseph, and upper Auglaize, the baseline phosphorus is the major contributor. According to Figures 7d and 7e, these regions have relatively low manure production and their fertilizer application rate is below the crop uptake rate.

Table 3 enumerates the phosphorus release mass by source type in 2019, totalling 4,057 tons of total phosphorus, with 46%, 26% and 29% from manure, fertilizer and baseline phosphorus, respectively. Overall, the manure contribution is higher than the fertilizer and baseline contributions in the basin, but the contributions vary substantially between different regions potentially due to differences in agricultural practices and manure production.

3.3 Phosphorus release peaks during spring planting period

Phosphorus transport from land to streams is driven by runoff, slope, soil condition, snow accumulation and crops (N. C. Hansen et al., 2000; Vadas et al., 2011; Zhang et al., 2019). Increased runoff accelerates phosphorus transport, and the transport can potentially increase many-fold if soil is loose and crop roots are short (Blanco-Canqui et al., 2004; Aronsson et al., 2016). Soil particles are generally agitated by precipitation

Table 3. Attribution of phosphorus to manure, fertilizer and base phosphorus. The attribution represents the outputs for the year 2019

Watershed Name	Area km^2	Total P tons	Manure P tons	Fertilizer P tons	Manure %	Fertilizer %	Baseline P %
Auglaize	4,316	1,612	721	363	45	23	33
St. Marys	2,054	1,199	660	211	55	18	27
St. Joseph	2,830	1,077	498	276	46	26	28
Tiffin	2,014	842	294	346	35	41	24
Upper Maumee	1,003	827	463	189	56	23	21
Blanchard	1,999	506	152	173	30	34	36
Maumee	13,969	4,057	1,847	1,037	46	26	29

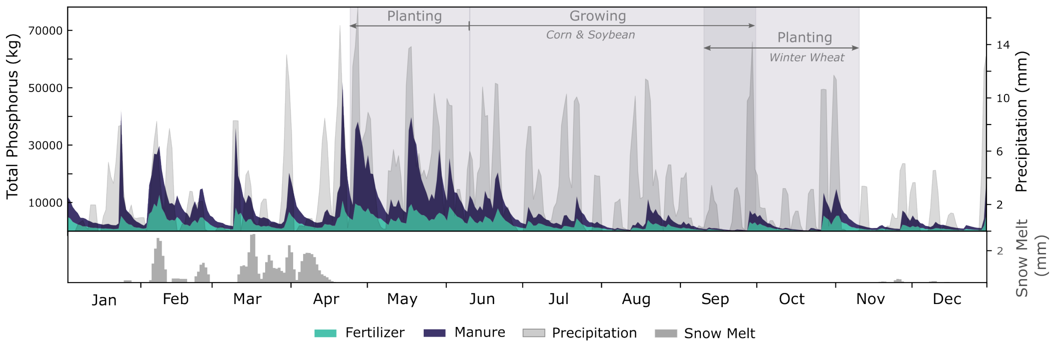

Figure 8. Daily total phosphorus mass in the streams at Maumee attributed to manure (purple) and fertilizer (green). Precipitation and snow melt time series, smoothed with a 3-days rolling mean, are shown with the light gray shaded area in the top panel and darker gray histogram in the bottom panel. Planting and growing periods for corn and soybeans as well as the planting period for winter wheat are depicted with with gray shaded rectangles. Note that the contribution from baseline phosphorus is not shown.

events, with intense precipitation making the land particularly vulnerable to erosion (Sharpley et al., 2008). However, soil agitation, and therefore phosphorus transport, is also a function of crop type and growing stage (Gao et al., 2009; Guo et al., 2019). Crops with larger canopy and widespread root distribution have the ability to reduce soil agitation and hold the soil particles, reducing phosphorus movement compared to non-vegetative area (Reubens et al., 2007; Zuazo & Pleguezuelo, 2009).

Figure 8 shows the manure and fertilizer release time series at Maumee along with the precipitation, snow melt, as well as crop planting and growing periods. We have extracted the precipitation and snow melt data from DAYMET (Thornton et al., 2016) and display the 3-day rolling mean of these time-series. We estimate snow melt by computing the first-order difference in snow water equivalent between consecutive time steps. We highlight the difference between different crop stages by shading the planting and growing periods of important crops in Ohio and Indiana. The spring planting period for corn and soybean is April 24–June 10 (USDA Statistical Reporting Service, 1984). The growing periods of corn and soybean is July to October (USDA Statistical Reporting Service, 1984; Kast, 2018). The winter wheat planting period is October 1 to November 1 (USDA Statistical Reporting Service, 1984).

Figure 8 demonstrates that manure and fertilizer phosphorus transport are highest during the spring planting season. This finding can be attributed to three factors. First, frequent and high precipitation increases flow and soil agitation that enhances phosphorus mobility. Second, fertilizer and manure application during the spring planting time means that plenty of phosphorus is available for transport. Third, the underdeveloped roots of newly planted crops have limited ability of retaining soil, resulting in relatively high mobility of soil particles, especially without cover crops. Overall, our model results suggest that manure phosphorus release during the spring planting period is around one-third of the annual manure phosphorus. Figure 8 also shows total phosphorus is lower during the growing season (July–Oct). While precipitation events during growing time tend to be similar to those during the planting period, phosphorus availability is lower later in the year because of increased soil retention by developed root systems. Additionally, phosphorus availability near the surface has decreased due to crop uptake and movement to relatively deeper soil layers.

Snow accumulation and melt control phosphorus transport during the winter months, December through March. At Maumee, most precipitation during this period falls as snow that accumulates over the soil, with several rainfall events leading to melt (Figure 8). During the winter months, the overall phosphorus release is relatively low, with manure and fertilizer phosphorus applied during antecedent wheat planting and earlier time covered by snow. Several high phosphorus release events coincide with the snow melt events during February to April (Figure 8). Snow melt events expose covered phosphorus from earlier fertilizer and manure application and convey it into the stream, possibly along with manure that might have been applied illegally over snow during the antecedent winter (Lewis & Makarewicz, 2009).

3.4 Additional upstream water quality monitoring reduces ambiguity in source attribution

In practice, it requires significant cost and effort to deploy water quality monitors in a watershed for pollution source attribution or to add new stations to an existing monitor network. Therefore, to maximize the useful information we can acquire from the limited monitors, we must be strategic about their placement locations. In this section, we use a leave-one-out cross validation analysis to first quantitatively demonstrate the reduced ambiguity in source attribution by incorporating the current water quality measurements and benefit of additional monitors. Then we gain insights about optimal locations of additional monitors by comparing information gain from each monitor.

In the leave-one-out cross validation, we test how well estimates at a particular monitor node align with the ground truth observations when the model does not have access to these observations during the fitting procedure. Given a set of monitor nodes Q in a network, we run a set of $|Q|$ simulations such that for simulation $q \in Q$, monitor node q is not included as a target in the ABC algorithm. In the analysis, we compare the priors, the posteriors of the leave-one-out simulations, the posterior estimates of the full simulations, and the observations. As discussed previously, even when the model does have full access to the data from all nodes, the error between the simulated mass and the target mass is nonzero. Therefore, we analyze outputs from the full model as well for comparison.

By comparing the posteriors of the leave-one-out simulations, the posterior estimates of the full simulations, and the observations, we demonstrate the reduced ambiguity with additional monitors. By comparing the posteriors of the leave-one-out simulations with the priors, we demonstrate that the model is learning important generalizable information about the system dynamics from the data for regions without monitors too, instead of merely memorizing the target time series. Then we study the sensitivity of attribution results to particular nodes to quantify the relative importance of

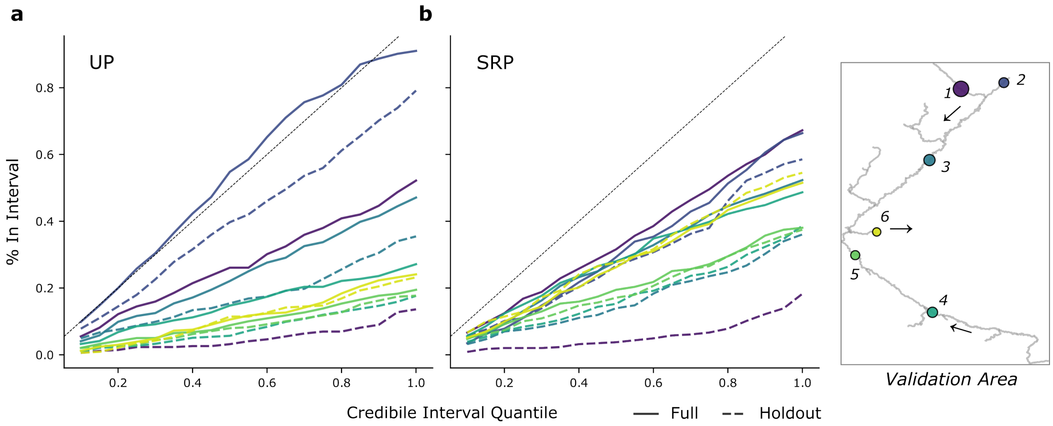

Figure 9. Evaluation of posteriors in validation study area (St. Marys and St. Joseph adjoined by outlet): Percentage of days where observed measurement falls within the X-% credibility interval plotted against the size of the credibility interval for (a) UP and (b) SRP. Each line represents the performance for a particular monitor node when the model has full access to all node data (solid line) and where the given node is held out (dashed). The color of the line corresponds to the monitor node in the map legend in the right panel, where the node size is proportional to mean Δ Full for UP and SRP listed in Table 4 representing information gain from the monitors.

each monitor node location, shedding light on the areas where additional monitors may be most beneficial.

Figure 9 visualizes the quality of the posteriors, in particular, the frequency with which the posteriors at a given monitor include the observed value. Each line depicts the proportion of days in which the observation falls within a given size credible interval as we vary the size of the interval (e.g., the .6 credible interval is the domain between the .2 and .8 quantiles). Each colored line corresponds to the posterior coverage at a particular monitor, with simulations when the given node is held out and included shown by the dotted and solid lines respectively. The thin dotted black line represents perfect coverage, that is, size of the credibility interval and coverage proportion are equal. Due to computational constraints, we only validate the model in a the western portion of the basin, shown in the validation area map in Figure 9. The validation area includes St. Marys and St. Joseph connected by the immediate downstream node draining to the Maumee River and the rest of the network.

Table 4 summarises the posterior coverage and provides point estimate errors for each held out node. We summarize the overall coverage of the posteriors by multiplying the total area under each curve in Figure 9 by 2; a coverage of 1 thus represents perfect coverage. The error is the same relative ℓ_1 distance used in ABC to compare the simulated and target mass (see Eq. (1)), where an error of 1 represents 100% difference in the estimate relative to the observed value. We also provide the error reduced or the coverage gained when the model has access to the data at the given monitor, which give an approximation of the relative importance of each monitor. The difference in the error of the leave-one-out run compared to the prior estimate (Δ Prior) represents performance gain at regions without monitor nodes from integrating real time water quality data into the model.

The posterior evaluation in Figure 9 and the summary in Table 4 reveal that the network model shows significant improvement over the prior, with the errors reduced by

Table 4. Validation metrics enumerated for the six monitors in the validation area, as depicted in the right panel of Figure 9. Error refers to the relative ℓ_1 error (see Eq. (1)) between the estimated and observed time series at the held out node, and Coverage is the sum of the area under the credible interval curves (multiplied by 2) shown in Figure 9. For each metric, we provide the value for each leave-one-out run (LOO), as well as the difference in each metric when the model has access to the observations at the given monitor node (Δ Full), and the difference in the metric compared to the prior estimate (Δ Prior). Positive difference means the metric for the leave-one-out run is lower. Note that we only provide the difference with the prior for the error metric, as the excess phosphorus method provides only a point estimate so that the coverage cannot be computed.

Monitor		**Error**			**Coverage**	
		LOO	Δ Full	Δ Prior	LOO	Δ Full
UP	1	0.182	-0.158	0.627	0.095	0.441
	2	0.057	-0.042	0.047	0.847	0.278
	3	0.072	-0.021	0.364	0.340	0.135
	4	0.085	-0.042	-0.005	0.162	0.164
	5	0.089	-0.003	0.030	0.161	0.027
	6	0.070	0.009	0.072	0.241	-0.016
	All	*0.093*	*-0.043*	*0.189*	*0.308*	*0.172*
SRP	1	0.316	-0.289	0.031	0.116	0.603
	2	0.108	-0.023	-0.054	0.611	0.032
	3	0.183	-0.119	-0.027	0.322	0.299
	4	0.085	-0.044	0.003	0.378	0.216
	5	0.082	-0.021	0.047	0.388	0.041
	6	0.057	0.004	0.076	0.602	0.002
	All	*0.139*	*-0.082*	*0.013*	*0.403*	*0.199*

82% and 63% on average for UP and SRP respectively from the priors to the posteriors of the full runs. It is notable that in the leave-one-out runs, the errors at the held out nodes still significantly decrease compared with the priors, with a mean reduction of 67% and 9% for UP and SRP, respectively. The improvements in the two comparisons demonstrate that learning from water quality measurements results in more accurate attribution throughout the stream network, rather than just at locations with monitors.

However, the importance of monitors, as measured by the change in error and coverage, varies significantly between monitors. The information gained by monitor 1 is particularly noticeable in the slope increase between the dashed and solid purple lines in Figure 9, with an mean coverage gain of 0.522 and error decrease of 0.223 across UP and SRP. We note that this particular monitor also demonstrates high relative updates, as shown by the dark red points well below the dotted line in Figure 5. On the other hand, the downstream-most monitor appears to provide no useful additional information. In general, the estimates are significantly more sensitive to the loss of data upstream than downstream, indicating that expanding monitoring upstream may be more beneficial for disambiguating sources of phosphorus pollution.

Note that even given the target monitor data during the prior update, the model estimates often deviate significantly from the ground truth. In fact, only the UP estimate at node 2 achieve near zero error (0.015) and perfect coverage (see dark blue solid line in Figure 9a), with other estimates falling well below this mark, averaging 0.487 coverage and 0.057 error for both forms of phosphorus. The high performance at monitor

2 is likely due to the fact very few subwatersheds lie above this monitor node and their priors generally have small updates, thus allowing the model to fit the observed data almost perfectly.

4 Discussion

Attributing sources of phosphorus has been a longstanding challenge at Maumee. High-resolution land use data (Boryan et al., 2011) and detailed data on manure production from CAFOs (EWG, 2019) have enabled public agencies like Environmental Working Group (EWG) and Environmental Law and Policy Center (ELPC) to map excess phosphorus over watersheds using a nutrient balance approach (ELPC, 2014; EWG, 2021). The resultant excess phosphorus estimation with high spatial resolution substantially advances identification of high-pollution areas and draws public attention to the problem of excessive agricultural phosphorus input. However, as shown in Figures 5 and 6, we found that equating such estimates with phosphorus losses to surface water can be inconsistent with water quality measurements (USGS, 2016; NCWQR, 2022).

While data of greater quality and quantity, such as detailed manure application ranges and finer fertilizer application data, can improve this nutrient balance approach, its fundamental limitation is the missing process connecting phosphorus input and loss. This process integrates factors like the spatiotemporal variations in runoff intensity, specific agricultural practices, and the biogeochemical evolution of phosphorus forms that are beyond the scope of a simple nutrient balance. Resolving these complexities in the style of modern hydrological models (Bicknell et al., 1993; Borah et al., 2002; Schwarz et al., 2006; Gironás et al., 2010; Arnold et al., 2012; Kast et al., 2019) would make source attribution expensive and inefficient. However, the resultant phosphorus loss, after being transported throughout the watershed, is recorded by water-quality measurements (USGS, 2016; NCWQR, 2022), which provide opportunities for effective and efficient attribution.

By integrating basic hydrological routing, our network model achieves greater accuracy than existing, data-based estimates of excess phosphorus (e.g., Figure 6). It leverages excess phosphorus estimates as a prior, integrates flow dynamics, and updates the prior by learning from water quality measurements. This updating process removes some of the bias of excess phosphorus in representing phosphorus loss, perhaps most importantly the tendency to overestimate pollution (Figure 5). Compared with the annual-scale estimates of excess phosphorus, our results reveal the temporal variation in phosphorus contribution, such as the immense contribution during spring planting and significant loss associated with snow melt (Figure 8).

Furthermore, using Approximate Bayesian Computation (ABC) without the need to define and evaluate likelihood functions (Beaumont et al., 2002; Csilléry et al., 2010; Sunnåker et al., 2013), our model is more lightweight with fewer parameters, as well as easier to set up and faster to run, than hydrologic models. Fine-scale source attribution with sparse monitors is an underdetermined problem. Using a probabilistic approach like ABC that generates posterior distributions has great advantage over deterministic approaches by covering possible scenarios and thus reducing the result bias. Although we use a beta prime distribution constructed based on excess phosphorus, the prior distribution for our model framework is flexible based on data availability and specific purposes, making our model framework suitable for application in other watersheds. In the ABC step of this study, we use the simple random sampling scheme, of which the required amount of samples quickly increases with the number of sources. Future work on implementing more advanced sampling scheme can potentially increase the efficiency and scalability of the model framework.

Our model framework may prove useful for policymakers and regulatory agencies seeking to make decisions about which pollutant sources to regulate, as well as how to write the rules governing these contributors. Given the limited resources of public agencies responsible for enforcement, like the U.S. Environmental Protection Agency (EPA), as well as the dearth of high-cadence water quality monitors, our model framework can also augment permitting and enforcement capacity by enabling agencies to focus scarce resources on facilities posing the highest risk. Our model enables spatial, temporal, and source-specific targeting of the most significant contributors without having to purchase and manage large computational resources or conduct labor-intensive monitoring. The model inferences, such as the high contribution of manure from upper St. Marys during spring planting, can enable evidence-based decisions regarding efficient resource allocation for pollution control. However, application in streams with significant phosphorus decay, such as the Lower Maumee River with its significant algal blooms, requires future work on modeling phosphorus sinks to release the current assumption on mass conservation.

Because adding new monitors to a stream network is costly, it requires evaluation of potential locations to maximize the benefit of additional monitors in attributing pollution to sources. Our model can help narrow down potential locations by quantifying the information gain from different monitors, as illustrated in 3.4. For Maumee, our results show that adding monitors to the upstream portion of watersheds, such as the upper St. Marys and upper St. Joseph, is the most beneficial for reducing ambiguity in source attribution (Figure 9, Table 4), because the further downstream the measurement, the larger the aggregated contribution from upstream regions reflected in it. The downstream region of the basin is often an area of concern, because the aggregated pollution from upstream leads to serious eutrophication, but our analysis suggests that the focus of water-quality monitoring needs to include, or even focus on, the upstream portion of the watershed.

At Maumee, the Clean Water Act efforts during the past years have resulted in improved nutrient management and decreases in the excessive soil phosphorus levels in some counties (Dayton et al., 2020). However, our results suggest baseline soil phosphorus remains a large contributor (Figure 7). These estimates remain highly uncertain, as we assume the baseline concentrations to be constant values, taken from runoff experiments (Sharpley, 1997). More accurate estimation requires relaxing this assumption by incorporating nutrient concentrations of runoff from cropland without recent fertilizer and manure application. Nonetheless, a high contribution from baseline soil phosphorus may still be expected given the high soil phosphorus levels in the MRB. For example, according to Dayton et al. (2020), the median Mehlich-3 soil test phosphorus (STP) levels of most counties still exceed 27 mg/kg, the upper bound of optimum (Dodd & Mallarino, 2005), with the larger quantiles of all counties greatly exceeding the optimum (Dayton et al., 2020). When STP exceeds optimum, the amount of phosphorus released from soil to runoff increases exponentially, leading to high phosphorus concentration even without additional fertilization (Higgs et al., 2000; Kleinman et al., 2002; Weil & Brady, 2017).

Increasing surplus phosphorus resultant from imbalanced input and output has been a global problem in developed and emerging economies (Bouwman et al., 2013). Besides causing direct phosphorus loss into aquatic systems (Figures 7a and 7b), high surplus phosphorus also accumulates in agricultural soils and leads to high baseline soil phosphorus levels (Weil & Brady, 2017). As exemplified by the case of Maumee in this study, baseline soil phosphorus from agricultural land is a large nonpoint source of pollution. Therefore, reducing excessive soil phosphorus and reducing its loss from agricultural land is crucial for nutrient management. Under our current model framework, the specific sources of the baseline soil phosphorus are unattributable, and its accumulation is a result of long-term fertilization (Nair et al., 1995).

One way of mitigating excessive soil phosphorus is to reduce fertilization on cultivated cropland (Sheffield et al., 2008). This reduction can be achieved via direct halting or reduction of fertilizer and manure application (McDowell et al., 2020), or phosphorus removal from manure (Lorimor et al., 2000; Sheffield et al., 2008). Another way is increasing plant uptake via double cropping of corn and winter cereals (Sheffield et al., 2008). Practices that reduce phosphorus loss from high-phosphorus soils include planting riparian buffers or cover crops, which reduce runoff intensity and absorb nutrients (Zhou et al., 2014; Weil & Brady, 2017).

5 Conclusions

This study advances our ability to attribute phosphorus sources by developing a lightweight modeling framework that integrates excess phosphorus derived from data, flow dynamics derived from hydrologic model, and water quality measurements data into a network model framework and applies the statistical approach Approximate Bayesian Computation. Our model reveals significant spatial and temporal variability in phosphorus release, which is averaged out in the coarse-scale attribution by calculating the difference between nutrient load measurements at sparsely deployed monitors. Being able to identify such variability can benefit targeted enforcement via prioritizing regions and time periods with higher pollutant release.

Open Research Section

v1.0.1 of the code used for the network model framework (Verma, Wei, et al., 2022) is preserved at `https://doi.org/10.5281/zenodo.7246383` with open access. The usage instructions are provided in the README files of the repository. All the processed data used in the simulation, part of the raw data, and the SWAT simulation results used by the network model framework (Verma, Alam, et al., 2022) are preserved at `https://doi.org/10.5281/zenodo.7295662` with open access. The code for processing the raw data, which are either in the data repository or publicly available online, is provided in the code repository. The links to the publicly available raw data are also provided in the code repository.

Acknowledgments
This work is supported by Stanford Impact Labs, Stanford Woods Institute for the Environment, The Chicago Community Trust, and Chicago Community Foundation, and by the Stanford Graduate Fellowship in Science and Engineering awarded to ZW.

References

Abbaspour, K. C., Rouholahnejad, E., Vaghefi, S., Srinivasan, R., Yang, H., & Kløve, B. (2015). A continental-scale hydrology and water quality model for Europe: Calibration and uncertainty of a high-resolution large-scale SWAT model. *Journal of hydrology*, *524*, 733–752.

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.

Arnold, J. G., Srinivasan, R., Muttiah, R. S., & Williams, J. R. (1998). Large area hydrologic modeling and assessment part i: model development 1. *Journal of the American Water Resources Association*, *34*(1), 73–89.

Aronsson, H., Hansen, E., Thomsen, I. K., Liu, J., Øgaard, A., Känkänen, H., & Ulén, B. (2016). The ability of cover crops to reduce nitrogen and phosphorus losses from arable land in southern Scandinavia and Finland. *Journal of Soil and water Conservation*, *71*(1), 41–55.

Baker, L. A. (1992). Introduction to nonpoint source pollution in the United States and prospects for wetland use. *Ecological Engineering*, *1*(1-2), 1–26.

Barnett, G. (1994). Phosphorus forms in animal manure. *Bioresource Technology*, *49*(2), 139-147. Retrieved from `https://www.sciencedirect.com/science/article/pii/0960852494900779` doi: https://doi.org/10.1016/0960-8524(94)90077-9

Beaumont, M. A., Zhang, W., & Balding, D. J. (2002). Approximate Bayesian Computation in population genetics. *Genetics*, *162*(4), 2025-2035. Retrieved from `https://doi.org/10.1093/genetics/162.4.2025` doi: 10.1093/genetics/162.4.2025

Bertol, O. J., Rizzi, N. E., Favaretto, N., & Lana, M. d. C. (2010). Phosphorus loss by surface runoff in no-till system under mineral and organic fertilization. *Scientia Agricola*, *67*(1), 71–77.

Bicknell, B., Imhoff, J., Kittle, J., Donigian, A., & Johanson, R. (1993). *Hydrologic Simulation Program—FORTRAN (HSPF) user's manual for Release 10. Report No. EPA/600/R-93/174.* US EPA Environmental Research Laboratory Athens, GA.

Bingham, M., Sinha, S., & Lupi, F. (2015). *Economic benefits of reducing harmful algal blooms in Lake Erie* (Tech. Rep.). Environmental Consulting & Technology, Inc.

Blanco-Canqui, H., Gantzer, C., Anderson, S., Alberts, E., & Thompson, A. (2004). Grass barrier and vegetative filter strip effectiveness in reducing runoff, sediment, nitrogen, and phosphorus loss. *Soil Science Society of America Journal*, *68*(5), 1670–1678.

Borah, D. K., Xia, R., Bera, M., et al. (2002). DWSM-a dynamic watershed simulation model. In V. P. Singh & D. K. Frevert (Eds.), *Mathematical models of small watershed hydrology and applications* (p. 113-166). Highlands Ranch, Colorado: Water Resources Publications, LLC.

Boryan, C., Yang, Z., Mueller, R., & Craig, M. (2011). Monitoring US agriculture: the US Department of Agriculture, National Agricultural Statistics Service, Cropland Data Layer Program. *Geocarto International*, *26*(5), 341–358.

Botts, L., & Muldoon, P. (2005). *Evolution of the Great Lakes water quality agreement.* Michigan State University Press.

Bouwman, L., Goldewijk, K. K., Hoek, K. W. V. D., Beusen, A. H. W., Vuuren, D. P. V., Willems, J., et al. (2013). Exploring global changes in nitrogen and phosphorus cycles in agriculture induced by livestock production over the 1900-2050 period. *Proceedings of the National Academy of Sciences*, *110*(52), 20882-20887. Retrieved from `https://www.pnas.org/doi/abs/10.1073/pnas.1012878108` doi: 10.1073/pnas.1012878108

Bridgeman, T. B., Chaffin, J. D., Kane, D. D., Conroy, J. D., Panek, S. E., & Armenio, P. M. (2012). From river to lake: Phosphorus partitioning and algal community compositional changes in Western Lake Erie. *Journal of Great Lakes Research*, *38*(1), 90–97.

Carpenter, S. R., Caraco, N. F., Correll, D. L., Howarth, R. W., Sharpley, A. N., & Smith, V. H. (1998). Nonpoint pollution of surface waters with phosphorus and nitrogen. *Ecological applications*, *8*(3), 559–568.

Chinowsky, P., Diekmann, J., & Galotti, V. (2008). Social network model of construction. *Journal of construction engineering and management*, *134*(10), 804–812.

Csilléry, K., Blum, M. G., Gaggiotti, O. E., & François, O. (2010). Approximate Bayesian Computation (ABC) in practice. *Trends in Ecology & Evolution*, *25*(7), 410-418. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0169534710000662` doi: https://doi.org/10.1016/j.tree.2010.04.001

Culman, S., Fulford, A., Camberato, J., & Steinke, K. (2020). *Tri-state fertilizer recommendations. bulletin 974.* College of Food, Agricultural, and Environmental Sciences. Columbus, OH: The Ohio State University.

Dayton, E. A., Shrestha, R. K., Fulford, A. M., Love, K. R., W. Culman, S., & Lindsey, L. E. (2020). Soil test phosphorus and phosphorus balance trends: A county-level analysis in Ohio. *Agronomy Journal*, *112*(3), 1617-1624. Retrieved from https://acsess.onlinelibrary.wiley.com/doi/abs/10.1002/agj2.20146 doi: https://doi.org/10.1002/agj2.20146

Dodd, J. R., & Mallarino, A. P. (2005). Soil-test phosphorus and crop grain yield responses to long-term phosphorus fertilization for corn-soybean rotations. *Soil Science Society of America Journal*, *69*(4), 1118-1128. Retrieved from https://acsess.onlinelibrary.wiley.com/doi/abs/10.2136/sssaj2004.0279 doi: https://doi.org/10.2136/sssaj2004.0279

Downing, J. A., Polasky, S., Olmstead, S. M., & Newbold, S. C. (2021). Protecting local water quality has global benefits. *Nature communications*, *12*(1), 1–6.

Easton, Z. M., Gérard-Marchant, P., Walter, M. T., Petrovic, A. M., & Steenhuis, T. S. (2007). Identifying dissolved phosphorus source areas and predicting transport from an urban watershed using distributed hydrologic modeling. *Water Resources Research*, *43*(11).

ELPC. (2014). *Using satellite imagery to count animal feeding operations* (Tech. Rep.). Environmental Law & Policy Center. Retrieved from https://elpc.org/wp-content/uploads/2020/05/CAFOFactsheet_ELPC_DRAFTv5.pdf

EWG. (2019). *Explosion of unregulated factory farms in Maumee Watershed fuels Lake Erie's toxic blooms* (Tech. Rep.). Environmental Working Group. Retrieved from https://www.ewg.org/interactive-maps/2019_maumee/

EWG. (2021). *Double trouble: Wisconsin's land and water are inundated with pollution from animal manure and excess farm fertilizer* (Tech. Rep.). Environmental Working Group. Retrieved from https://www.ewg.org/node/7173/

EWG. (2022). *News reports of algae blooms, 2010 to present.* Retrieved from https://www.ewg.org/interactive-maps/algal_blooms/map/

Falcone, J. A. (2021). *Estimates of county-level nitrogen and phosphorus from fertilizer and manure from 1950 through 2017 in the conterminous United States* (Tech. Rep.). US Geological Survey.

Fordham, A. W., & Schwertmann, U. (1977). Composition and reactions of liquid manure (gülle), with particular reference to phosphate: I. analytical composition and reaction with poorly crystalline iron oxide (ferrihydrite). *Journal of Environmental Quality*, *6*(2), 133-136. Retrieved from https://acsess.onlinelibrary.wiley.com/doi/abs/10.2134/jeq1977.00472425000600020006x doi: https://doi.org/10.2134/jeq1977.00472425000600020006x

Gao, Y., Zhu, B., Zhou, P., Tang, J.-L., Wang, T., & Miao, C.-Y. (2009). Effects of vegetation cover on phosphorus loss from a hillslope cropland of purple soil under simulated rainfall: a case study in China. *Nutrient Cycling in Agroecosystems*, *85*(3), 263–273.

Gironás, J., Roesner, L. A., Rossman, L. A., & Davis, J. (2010). A new applications manual for the storm water management model(swmm). *Environmental Modelling & Software*, *25*(6), 813–814.

Guo, M., Zhang, T., Li, J., Li, Z., Xu, G., & Yang, R. (2019). Reducing nitrogen and phosphorus losses from different crop types in the water source area of the danjiang river, china. *International Journal of Environmental Research and Public Health*, *16*(18), 3442.

Gupta, H. V., Sorooshian, S., & Yapo, P. O. (1998). Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information. *Water Resources Research*, *34*(4), 751–763.

Han, W., Yang, Z., Di, L., & Mueller, R. (2012). CropScape: A Web service based application for exploring and disseminating US conterminous geospatial cropland data products for decision support. *Computers and Electronics in Agriculture*, *84*, 111–123.

Hansen, J. C., Cade-Menun, B. J., & Strawn, D. G. (2004). Phosphorus speciation in manure-amended alkaline soils. *Journal of Environmental Quality*, *33*(4), 1521-1527. Retrieved from `https://acsess.onlinelibrary.wiley.com/doi/abs/10.2134/jeq2004.1521` doi: https://doi.org/10.2134/jeq2004.1521

Hansen, N. C., Gupta, S. C., & Moncrief, J. F. (2000). Snowmelt runoff, sediment, and phosphorus losses under three different tillage systems. *Soil and tillage research*, *57*(1-2), 93–100.

Hart, M. R., Quin, B. F., & Nguyen, M. L. (2004). Phosphorus runoff from agricultural land and direct fertilizer effects: A review. *Journal of environmental quality*, *33*(6), 1954–1972.

Higgs, B., Johnston, A., Salter, J., & Dawson, C. (2000). Some aspects of achieving sustainable phosphorus use in agriculture. *Journal of Environmental Quality*, *29*(1), 80–87.

Howarth, R. W., Anderson, D., Cloern, J. E., Elfring, C., Hopkinson, C. S., Lapointe, B., et al. (2000). Nutrient pollution of coastal rivers, bays, and seas. *Issues in ecology*(7), 1–16.

International Joint Commission. (1978). *Great Lakes water quality agreement of 1978: Agreement, with annexes and terms of reference, between the United States of America and Canada, signed at Ottawa, November 22, 1978.* International Joint Commission.

Kane, D. D., Conroy, J. D., Richards, R. P., Baker, D. B., & Culver, D. A. (2014). Re-eutrophication of Lake Erie: Correlations between tributary nutrient loads and phytoplankton biomass. *Journal of Great Lakes Research*, *40*(3), 496–501.

Kast, J. B. (2018). *Manure management in the Maumee River Watershed and watershed modeling to assess impacts on Lake Erie's water quality* (Unpublished doctoral dissertation). The Ohio State University.

Kast, J. B., Long, C. M., Muenich, R. L., Martin, J. F., & Kalcic, M. M. (2019). Manure management at ohio confined animal feeding facilities in the maumee river watershed. *Journal of Great Lakes Research*, *45*(6), 1162–1170.

Keiser, D. A., & Shapiro, J. S. (2019). Us water pollution regulation over the past half century: burning waters to crystal springs? *Journal of Economic Perspectives*, *33*(4), 51–75.

Khuller, S., & Raghavachari, B. (1996). Graph and network algorithms. *ACM Computing Surveys (CSUR)*, *28*(1), 43–45.

Kleinman, P. J. A., Sharpley, A. N., Moyer, B. G., & Elwinger, G. F. (2002). Effect of mineral and manure phosphorus sources on runoff phosphorus. *Journal of Environmental Quality*, *31*(6), 2026-2033. Retrieved from `https://acsess.onlinelibrary.wiley.com/doi/abs/10.2134/jeq2002.2026` doi: https://doi.org/10.2134/jeq2002.2026

Kling, H., Fuchs, M., & Paulin, M. (2012). Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *Journal of Hydrology*, *424*, 264–277.

Kumar, R. R., Park, B. J., & Cho, J. Y. (2013). Application and environmental risks of livestock manure. *Journal of the Korean Society for Applied Biological Chemistry*, *56*(5), 497–503.

Lewis, T. W., & Makarewicz, J. C. (2009). Winter application of manure on an agricultural watershed and its impact on downstream nutrient fluxes. *Journal of Great Lakes Research*, *35*, 43-49. Retrieved from `https://www.sciencedirect.com/science/article/pii/S038013300900046X` (Special issue on Watershed Management and Nearshore Lake Water Quality, The Conesus Lake Watershed Study) doi: https://doi.org/10.1016/j.jglr.2008.08.003

Lin, F.-R., Wu, N.-J., Tu, C.-H., & Tsay, T.-K. (2017). Automatic calibration of an unsteady river flow model by using dynamically dimensioned search algorithm. *Mathematical Problems in Engineering*, *2017*.

Liu, Y., Li, H., Cui, G., & Cao, Y. (2020). Water quality attribution and simulation of non-point source pollution load flux in the Hulan River basin. *Scientific Reports*, *10*(1), 1–15.

Long, C. M., Muenich, R. L., Kalcic, M. M., & Scavia, D. (2018). Use of manure nutrients from concentrated animal feeding operations. *Journal of Great Lakes Research*, *44*(2), 245-252. Retrieved from https://www.sciencedirect.com/science/article/pii/S0380133018300078 doi: https://doi.org/10.1016/j.jglr.2018.01.006

Lorimor, J., Powers, W., & Sutton, A. (2000). *Manure characteristics.* Iowa State University. Retrieved from https://www.canr.msu.edu/uploads/files/ManureCharacteristicsMWPS-18_1.pdf

Matott, L. (2017). *OSTRICH: an optimization software tool, documentation and user's guide, Version 17.12.19.* (Tech. Rep.). University at Buffalo Center for Computational Research. Retrieved from http://www.civil.uwaterloo.ca/envmodelling/Ostrich.html

McDowell, R., Dodd, R., Pletnyakov, P., & Noble, A. (2020). The ability to reduce soil legacy phosphorus at a country scale. *Frontiers in Environmental Science*, *8*. Retrieved from https://www.frontiersin.org/articles/10.3389/fenvs.2020.00006 doi: 10.3389/fenvs.2020.00006

Michalak, A. M., Anderson, E. J., Beletsky, D., Boland, S., Bosch, N. S., Bridgeman, T. B., et al. (2013). Record-setting algal bloom in Lake Erie caused by agricultural and meteorological trends consistent with expected future conditions. *Proceedings of the National Academy of Sciences*, *110*(16), 6448–6452.

Mohamed, M. N., Wellen, C., Parsons, C. T., Taylor, W. D., Arhonditsis, G., Chomicki, K. M., et al. (2019). Understanding and managing the re-eutrophication of Lake Erie: Knowledge gaps and research priorities. *Freshwater Science*, *38*(4), 675-691. Retrieved from https://doi.org/10.1086/705915 doi: 10.1086/705915

Myers, D. N., Metzker, K. D., & Davis, S. (2000). *Status and trends in suspended-sediment discharges, soil erosion, and conservation tillage in the Maumee River Basin–Ohio, Michigan, and Indiana.* Reston, VA: U.S. Geological Survey. Retrieved from https://doi.org/10.3133/wri004091 doi: 10.3133/wri004091

Nair, V. D., Graetz, D. A., & Portier, K. M. (1995). Forms of phosphorus in soil profiles from dairies of south Florida. *Soil Science Society of America Journal*, *59*(5), 1244-1249. Retrieved from https://acsess.onlinelibrary.wiley.com/doi/abs/10.2136/sssaj1995.03615995005900050006x doi: https://doi.org/10.2136/sssaj1995.03615995005900050006x

NCWQR. (2022). *National center for water quality research, sustaining our soil and water resources.* Retrieved from https://ncwqr.org/

OC Interagency WQI Workgroup. (2017). *Water quality indicators (WQI) project background and technical specifications* (Tech. Rep.). US EPA Office of Compliance. Retrieved from https://echo.epa.gov/tools/data-downloads/wqi-data-review

Ohio EPA. (2016). *Nutrient mass balance study for Ohio's major rivers* (Tech. Rep.). Columbus, OH: Ohio Environmental Protection Agency, Division of Surface Water, Modeling, Assessment and TMDL Section.

Ongley, E. D., Xiaolan, Z., & Tao, Y. (2010). Current status of agricultural and rural non-point source pollution assessment in China. *Environmental Pollution*, *158*(5), 1159–1168.

Parry, R. (1998). Agricultural phosphorus and water quality: A us environmental protection agency perspective. *Journal of Environmental Quality*, *27*(2), 258–261.

Pokorádi, L. (2018). Graph model-based analysis of technical systems. In *Iop conference series: Materials science and engineering* (Vol. 393, p. 012007).

PRISM Climate Group. (2014). *PRISM Gridded Climate Data.* (Oregon State University. https://prism.oregonstate.edu)

Ramos, M., & Martínez-Casasnovas, J. (2006). Nutrient losses by runoff in vineyards of the Mediterranean Alt Penedès region (NE Spain). *Agriculture, Ecosystems & Environment*, *113*(1), 356-363. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0167880905005153` doi: https://doi.org/10.1016/j.agee.2005.10.009

Reubens, B., Poesen, J., Danjon, F., Geudens, G., & Muys, B. (2007). The role of fine and coarse roots in shallow slope stability and soil erosion control with a focus on root system architecture: a review. *Trees*, *21*(4), 385–402.

Scavia, D., Allan, J. D., Arend, K. K., Bartell, S., Beletsky, D., Bosch, N. S., et al. (2014). Assessing and addressing the re-eutrophication of Lake Erie: Central basin hypoxia. *Journal of Great Lakes Research*, *40*(2), 226–246.

Schwarz, G. E., Hoos, A. B., Alexander, R., & Smith, R. (2006). *The SPARROW surface water-quality model: theory, application and user documentation* (Tech. Rep.). US Geological Survey.

Sharpley, A. N. (1995). Dependence of runoff phosphorus on extractable soil phosphorus. *Journal of Environmental Quality*, *24*(5), 920-926. Retrieved from `https://acsess.onlinelibrary.wiley.com/doi/abs/10.2134/jeq1995.00472425002400050020x` doi: https://doi.org/10.2134/jeq1995.00472425002400050020x

Sharpley, A. N. (1997). Rainfall frequency and nitrogen and phosphorus runoff from soil amended with poultry litter. *Journal of Environmental Quality*, *26*(4), 1127-1132. Retrieved from `https://acsess.onlinelibrary.wiley.com/doi/abs/10.2134/jeq1997.00472425002600040026x` doi: https://doi.org/10.2134/jeq1997.00472425002600040026x

Sharpley, A. N., Kleinman, P. J., Heathwaite, A. L., Gburek, W. J., Folmar, G. J., & Schmidt, J. P. (2008). Phosphorus loss from an agricultural watershed as a function of storm size. *Journal of environmental quality*, *37*(2), 362–368.

Sheffield, R., Brown, B., Chahine, M., de Haro Marti, M., & Falen, C. (2008). Mitigating high phosphorus soils. *University of Idaho Extension*(851). Retrieved from `https://www.uidaho.edu/extension/publications`

Shen, Z., Liao, Q., Hong, Q., & Gong, Y. (2012). An overview of research on agricultural non-point source pollution modelling in China. *Separation and Purification Technology*, *84*, 104–111.

Soil Survey Staff. (2015). *Soil Survey Geographic (SSURGO) Database for Maumee River Basin* (Tech. Rep.). Natural Resources Conservation Service, United States Department of Agriculture. Retrieved from `https://data.nal.usda.gov/dataset/soil-survey-geographic-database-ssurgo`

Stackpoole, S. M., Stets, E. G., & Sprague, L. A. (2019). Variable impacts of contemporary versus legacy agricultural phosphorus on US river water quality. *Proceedings of the National Academy of Sciences*, *116*(41), 20562–20567.

Stets, E. G., Sprague, L. A., Oelsner, G. P., Johnson, H. M., Murphy, J. C., Ryberg, K., et al. (2020). Landscape drivers of dynamic change in water quality of US rivers. *Environmental Science & Technology*, *54*(7), 4336–4343.

Stoddard, J. L., Van Sickle, J., Herlihy, A. T., Brahney, J., Paulsen, S., Peck, D. V., et al. (2016). Continental-scale increase in lake and stream phosphorus: Are oligotrophic systems disappearing in the united states? *Environmental Science & Technology*, *50*(7), 3409–3415.

Sunnåker, M., Busetto, A. G., Numminen, E., Corander, J., Foll, M., & Dessimoz, C. (2013). Approximate Bayesian Computation. *PLoS computational biology*, *9*(1), e1002803.

Thornton, P., Thornton, M., Mayer, B., Wei, Y., Devarakonda, R., Vose, R., & Cook, R. (2016). Daymet: daily surface weather data on a 1-km grid for north america, version 3. ornl daac, oak ridge, tennessee, usa. *USDA-NASS, 2019. 2017 Census of Agriculture, Summary and State Data, Geographic Area Series, Part 51, AC-17-A-51*.

Tolson, B. A., & Shoemaker, C. A. (2007). Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. *Water Resources Research*, *43*(1).
USDA Statistical Reporting Service. (1984). *Usual planting and harvesting dates for us field crops* (No. 628). US Department of Agriculture.
USDA-NASS. (2021). *2021 state agriculture overview* (Tech. Rep.). United States Department of Agriculture. Retrieved from `https://www.nass.usda.gov/Statistics_by_State/Ohio/Publications/Annual_Statistical_Bulletin/`
USEPA. (2003). National pollutant discharge elimination system permit regulation and effluent limitation guidelines and standards for concentrated animal feeding operations (cafos); final rule. *Fed. Regist.*, *68*(29), 7176–7274.
USEPA, & USGS. (2012). *NHDPlusV2 (National Hydrography Dataset Plus Version 2).* Retrieved from `https://www.epa.gov/waterdata/nhdplus-national-hydrography-dataset-plus`
USGS. (2016). *National Water Information System: USGS Water Data for USA.* US Geological Surve. Retrieved from `https://waterdata.usgs.gov/nwis`
Vadas, P. A., Jokela, W. E., Franklin, D. H., & Endale, D. M. (2011). The effect of rain and runoff when assessing timing of manure application and dissolved phosphorus loss in runoff 1. *Journal of the American Water Resources Association*, *47*(4), 877–886.
Verma, M., Alam, S., Wei, Z., Hilderbran, M., Wu, Y., Anderson, B., et al. (2022). *Data repository for "integrating water quality data with a bayesian network model to improve spatial and temporal phosphorus attribution: Application to the maumee river basin".* Zenodo. Retrieved from `https://doi.org/10.5281/zenodo.7295662` doi: 10.5281/zenodo.7295662
Verma, M., Wei, Z., Alam, S., Hilderbran, M., Wu, Y., Anderson, B., et al. (2022). *Code repository for "integrating water quality data with a bayesian network model to improve spatial and temporal phosphorus attribution: Application to the maumee river basin".* Zenodo. Retrieved from `https://doi.org/10.5281/zenodo.7246383` doi: 10.5281/zenodo.7246383
Watters, H. (2021). *Nutrient removal for field crops in Ohio* (Tech. Rep.). Columbus, Ohio: Ohio State University. Retrieved from `https://ohioline.osu.edu/factsheet/anr-96`
Weil, R., & Brady, N. (2017). *The nature and properties of soils. 15th edition.* Pearson Education.
Whitehead, P. G., Wilby, R. L., Battarbee, R. W., Kernan, M., & Wade, A. J. (2009). A review of the potential impacts of climate change on surface water quality. *Hydrological sciences journal*, *54*(1), 101–123.
Yang, X., Liu, Q., Fu, G., He, Y., Luo, X., & Zheng, Z. (2016). Spatiotemporal patterns and source attribution of nitrogen load in a river basin with complex pollution sources. *Water research*, *94*, 187–199.
Zambrano-Bigiarini, M., & Rojas, R. (2013). A model-independent Particle Swarm Optimisation software for model calibration. *Environmental Modelling & Software*, *43*, 5–25.
Zhang, R., Li, M., Yuan, X., & Pan, Z. (2019). Influence of rainfall intensity and slope on suspended solids and phosphorus losses in runoff. *Environmental Science and Pollution Research*, *26*(33), 33963–33975.
Zhou, X., Helmers, M. J., Asbjornsen, H., Kolka, R., Tomer, M. D., & Cruse, R. M. (2014). Nutrient removal by prairie filter strips in agricultural landscapes. *Journal of Soil and Water Conservation*, *69*(1), 54–64.
Zuazo, V. H. D., & Pleguezuelo, C. R. R. (2009). Soil-erosion and runoff prevention by plant covers: a review. *Sustainable agriculture*, 785–811.

AGU PUBLICATIONS

Water Resources Research

Supporting Information for

Integrating Water Quality Data with a Bayesian Network Model to Improve Spatial and Temporal Phosphorus Attribution: Application to the Maumee River Basin

Zihan Wei[1]*, Sarfaraz Alam[1,2*], Miki Verma[3*], Margaret Hilderbran[2], Yuchen Wu[4], Brandon Anderson[2], Daniel E. Ho[2†], Jenny Suckale[1†]

[*]Joint First Authors
[†]Equal co-supervision
[1]Department of Geophysics, Stanford University, Stanford, CA, USA
[2]Regulation, Evaluation and Governance Lab, Stanford University, Stanford, CA, USA
[3]Symbolic Systems Program, Stanford University, Stanford, CA, USA
[4]Department of Statistics, Stanford University, Stanford, CA, USA

Contents of this file

Introduction

This supporting information document presents text, figures, and tables that provide additional details about the methods and results outlined in the manuscript.

Text S1. Hydrologic model set-up.

We used the Soil and Water Assessment Tool (SWAT), a semi-distributed, physically based hydrologic model, to simulate the hydrologic processes in the Maumee River Basin. The SWAT model solves the water balance equation at its smallest calculation unit, known as a hydrologic response unit (HRU), to quantify water flux and changes in storage. Each HRU is determined from the unique combinations of land use, soil, and slope data. The key strength of the SWAT model is that it can represent the physical hydrologic processes and model agricultural and water management changes, all while being computationally faster than commonly used distributed hydrologic models like Variable Infiltration Capacity (VIC).

We used topography, soil, land use, and meteorological time series data to set up the SWAT model for the Maumee River Basin. Table S1 lists the type, source, and resolution of each data set used for the SWAT model. We followed four key steps for model development.

First, we delineated the watersheds using topographic data (via a digital elevation model, or DEM). The elevation data was used to compute flow direction and flow accumulation (i.e., the number of grids contributing flow to each grid). Streams generally have relatively higher flow accumulation value (or higher number of grids upstream contributing flow), which then used to separate stream networks. Based on the threshold area for flow accumulation, the stream network density was determined. Using smaller thresholds yielded denser networks. We tested several thresholds in an attempt to obtain stream network density resembling the USGS HUC-12 watersheds, ultimately using a threshold of 3000 ha, or 30 km^2. However, we note that the areas we obtained are not identical to HUC-12 watersheds.

Furthermore, to obtain simulated outputs at locations with USGS water quality and flow measurements, we added outlet points at these locations. In a few cases, USGS monitor locations are not exactly on the streamlines due to errors in delineated stream network locations. The typical approach in this scenario is to snap the monitor locations to the streamline, which we did for distances up to 100 m from the stream. We note that positioning outlets subdivides a subwatershed into two, which in some cases resulted in the creation of much smaller subwatersheds. Furthermore, we identified two channels in the NHDPlus stream network that are undirected cycles (loops independent of edge direction); these generally occur when there are bypasses or irrigation channels. Because we seek to aggregate the channel contributions at subwatershed scale, we collapsed these loops into single edges. In summary, the watershed delineation process yields subwatersheds with outlets located at the water quality monitors and stream junctions. For reference, these monitor and junction nodes are later used to simulate pollutant transport through the stream network, while the subwatersheds are used as source nodes.

Second, we used the land use map from USDA Cropland Data Layer (Han et al., 2012), the soil map from SSURGO (Soil Survey Staff, 2015), and slope information derived from DEM to determine HRUs. SWAT used these three datasets to find unique combinations of land parcels, which are defined to be the HRUs. All simulation in SWAT is first computed at the HRU level, then aggregated at the subwatershed level.

Third, we forced the model with temperature and precipitation data from PRISM (PRISM Climate Group, 2014) to simulate the model at daily time steps from 2014 through 2020. We used 2014 as the spinning period (or warming period, which is necessary for model stability), so the simulation output is available from 2015 to 2020.

Fourth, we calibrated the model using SWAT-simulated streamflow as the calibration variable and the USGS streamflow data as the 'observed' data. The objective function for calibration was to maximize Kling-Gupta Efficiency (KGE). Details about calibration and validation are provided in the following section (Text S2).

Figure S1 shows the elevation, land use, and soil maps used as inputs to the SWAT model.

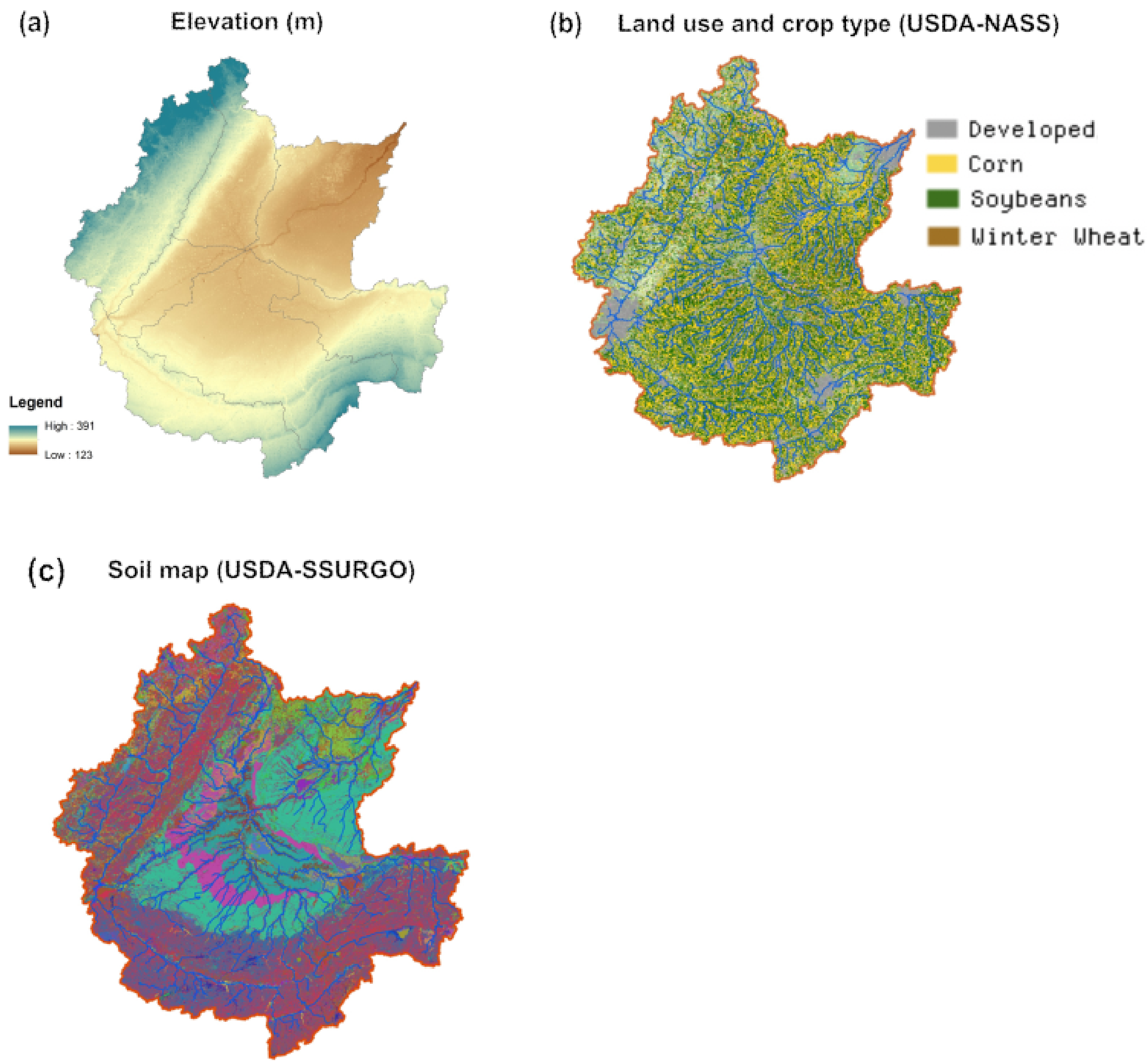

Figure S1. Input data to the SWAT hydrologic model. (a) 30-m elevation map from Shuttle Radar Topographic Mission (SRTM). (b) 30-m land use and crop type map from USDA-NASS. Legend for land use includes only the dominant land use types; others are not shown for concision. (c) 10-m SSURGO soil map from USDA. Legend for soil type, which consists of a large number of soil types, is not shown for brevity.

Text S2. SWAT calibration and validation.

When calibrating the SWAT model, we search for the model parameter values for which model simulation most closely matches the in-situ measurements. Because we use simulated flow and runoff for source attribution, we calibrate the model using streamflow, or river discharge. Hydrologic model calibration is generally suggested to be treated as a multi-objective problem using either multi-site or multi-variable measurements or multi-response function (Gupta et al., 2009; van Griensven & van Bauwens, 2003; Madsen, 2003). In this study we performed multi-site calibration. For calibration, we chose Kling-Gupta-Efficiency (KGE) as the objective function (Kling et al., 2012). KGE includes correlation (r), variability (α), and bias error (β) in its goodness-of-fit criterion (Gupta et al., 1998; Kling et al., 2012). This goodness-of-fit criterion measures the match between simulated and observed values on a scale ranging from negative infinity to 1, where 1 indicates a perfect match.

The objective function for optimization is:

$$KGE_{mean} = \sum_{j=1}^{k} \frac{1}{k} \left(1 - \sqrt{\left((r-1)^2 + (\alpha-1)^2 + (\beta-1)^2\right)_j}\right)$$

Where,

$$\alpha = \frac{\sigma_{sim}/\mu_{sim}}{\sigma_{obs}/\mu_{obs}}$$

and

$$\beta = \frac{\mu_{sim}}{\mu_{obs}}$$

and k is the total number of streamflow measurement sites; r is the regression coefficient, α is a measure of relative variability (variability ratio); and β is the bias ratio (the ratio of the simulated and observed means, μ_{sim} and μ_{obs}, respectively). σ is the standard deviation. We calibrated the streamflow for three years: 2015, 2017, and 2019. The validation periods were the alternate years: 2016, 2018, and 2020.

We used the Dynamically Dimensioned Search Algorithm (DDS), a widely used method for hydrologic calibration, to optimize SWAT model parameters (Lin et al., 2017; Tolson and Shoemaker, 2007). The key advantage of DDS over commonly-used global search algorithms (e.g., the shuffled complex evolution algorithm) is the ability to dynamically adjust search space by successively decreasing parameter dimension until iterations reach a user-defined limit. For this study, we calibrated seven parameters and iterated 3000 times, using the tool Ostrich that has a built-in DDS algorithm (Matott, 2017). Parameter selection was based on the most commonly-used parameters for streamflow calibration (Abbaspour et al., 2015; Zambrano-Bigiarini & Rojas, 2013), as well as our experiments to identify most sensitive parameters. The calibration parameters used in our SWAT model are listed in Table S1.

Table S1. Calibrated parameters for the SWAT model. Here, R indicates that an existing parameter value is multiplied by (1+ a given value), while V indicates that the existing parameter value is replaced by a given value.

Parameter	*Definition*	*Type of change*	*Range*	*Fitted value*
CN2	Curve number for moisture condition II	R	-0.25 _ 0.25	0.04
ALPHA_BF	Baseflow alpha factor for bank storage (days)	V	0 _ 1	0.99
SURLAG	Surface runoff lag coefficient	V	0.01 _ 2	0.3
GW_DELAY	Groundwater re-evaporation factor	V	0.01 _ 50	0.07
SOL_AWC	Available soil water capacity (mm H2O/mm soil)	R	-0.8 _ 0.2	0.04
SOIL_K	Saturated hydraulic conductivity (mm/h)	R	-0.8 _ 0.2	0.0017
GWQMN	Threshold depth of water in the shallow aquifer required for return flow to occur (mm)	V	-1000 _ 1000	657

Table S2 shows the performance metrics. We find the KGE values for the calibration and validation periods are 0.78 and 0.82, respectively. R^2 for the calibration and validation periods are 0.87 and 0.83, respectively. The KGE and R^2 values testing the match between simulated and observed flow indicate overall satisfactory model performance.

Table S2. Performance metrics for calibration and validation periods.

Evaluation criterion	*Calibration period*	*Validation period*
KGE	0.78	0.82
R^2	0.87	0.83

Figure S2 compares simulated and observed streamflows at multiple USGS sites.

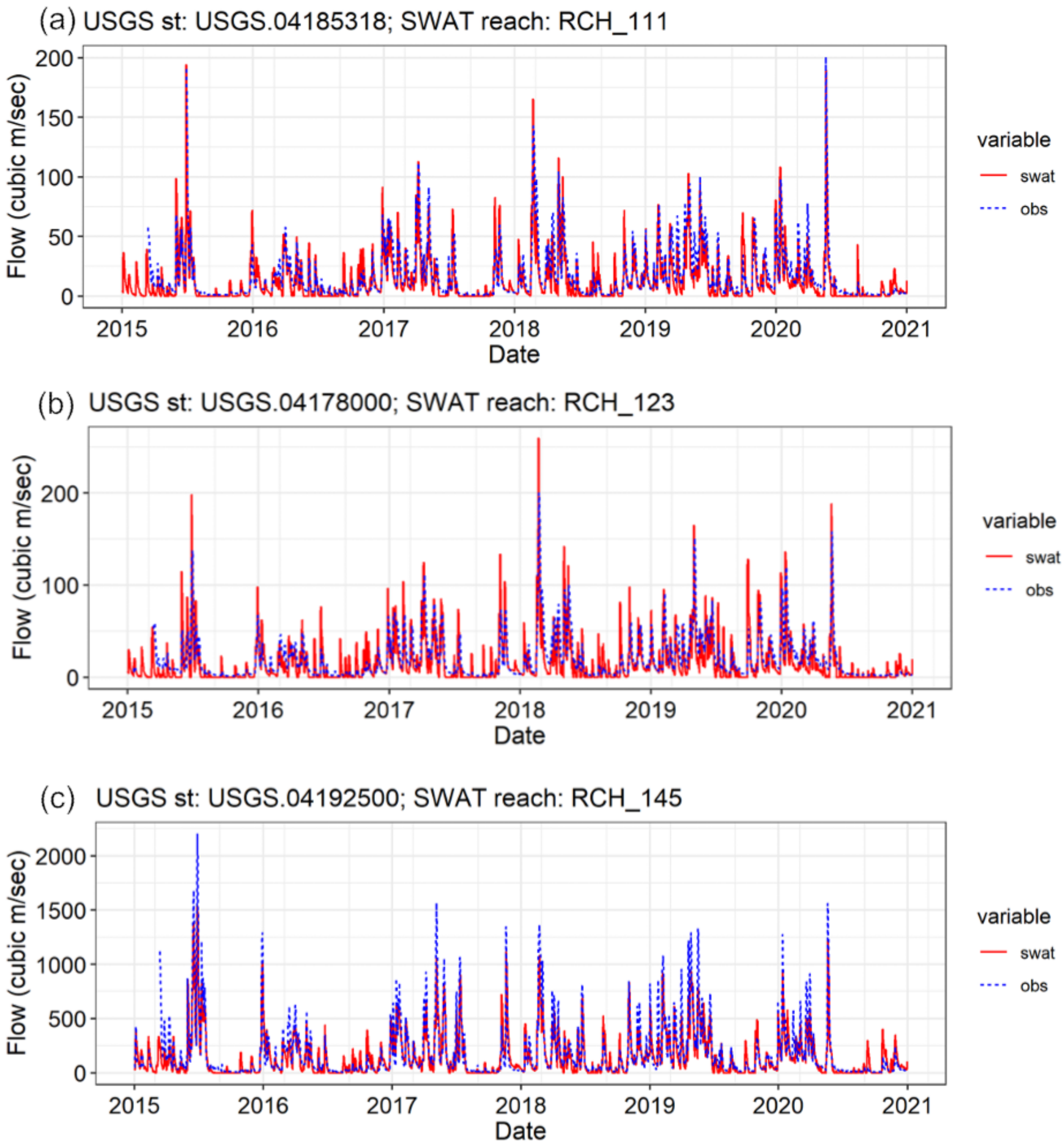

Figure S2. SWAT-simulated vs. USGS observed flow at selected sites.

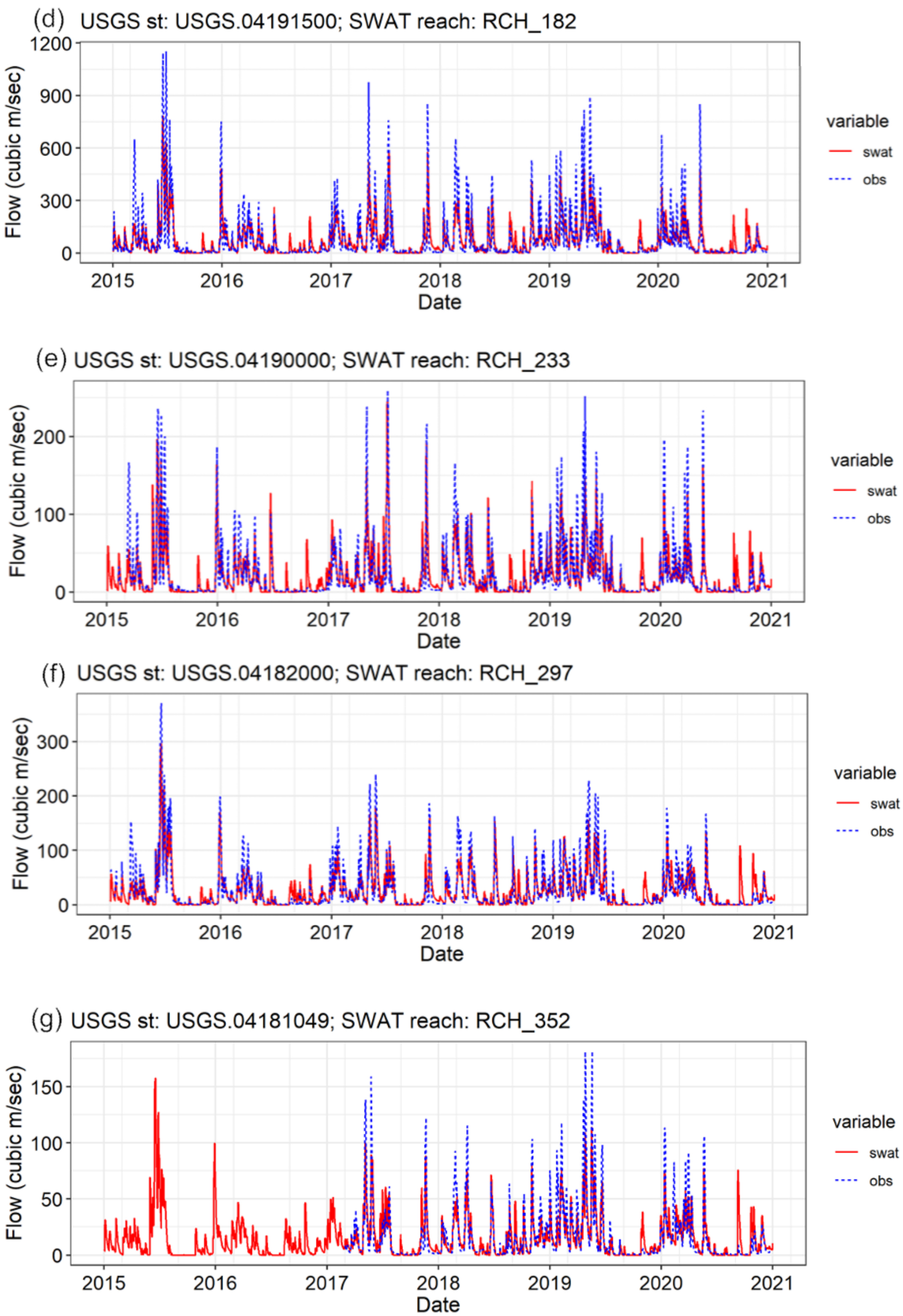

Figure S2. SWAT-simulated vs. USGS observed flow at selected sites (continued).

References

Abbaspour, K. C., Rouholahnejad, E., Vaghefi, S., Srinivasan, R., Yang, H. and Kløve, B. (2015). A continental-scale hydrology and water quality model for Europe: Calibration and uncertainty of a high-resolution large-scale SWAT model. *Journal of Hydrology*, *524*, 733-752. doi:10.1016/j.jhydrol.2015.03.027

Gupta, H. V., Sorooshian, S. and Yapo, P. O. (1998). Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information. *Water Resources Research*, *34*(4), 751-763. doi:10.1029/97WR03495

Han, W., Yang, Z., Di, L. and Mueller, R. (2012). CropScape: A Web service based application for exploring and disseminating US conterminous geospatial cropland data products for decision support. *Computers and Electronics in Agriculture*, *84*, 111-123. doi:10.1016/j.compag.2012.03.005

Kling, H., Fuchs, M. and Paulin, M. (2012). Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *Journal of Hydrology*, *424-425*, 264-277. doi:10.1016/j.jhydrol.2012.01.011

Lin, F. R., Wu, N. J., Tu, C. H. and Tsay, T. K. (2017). Automatic calibration of an unsteady river flow model by using dynamically dimensioned search algorithm. *Mathematical Problems in Engineering*, *2017*(7), 1-19. doi:10.1155/2017/7919324

Matott, L. S. (2017). *OSTRICH: an optimization software tool, documentation and user's guide, version 17.12.19*. University at Buffalo Center for Computational Research. http://www.civil.uwaterloo.ca/envmodelling/Ostrich.html.

PRISM Climate Group. (2014). PRISM Gridded Climate Data. Oregon State University. https://prism.oregonstate.edu

Soil Survey Staff. (2015). Soil Survey Geographic (SSURGO) Database for Maumee River Basin. Natural Resources Conservation Service, United States Department of Agriculture. https://data.nal.usda.gov/dataset/soil-survey-geographic-database-ssurgo

Tolson, B. A. and Shoemaker, C. A. (2007). Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. *Water Resources Research*, *43*(1). doi:10.1029/2005WR004723

Zambrano-Bigiarini, M. and Rojas, R. (2013). A model-independent Particle Swarm Optimisation software for model calibration. *Environmental Modelling & Software*, *43*, 5-25. doi:10.1016/j.envsoft.2013.01.004