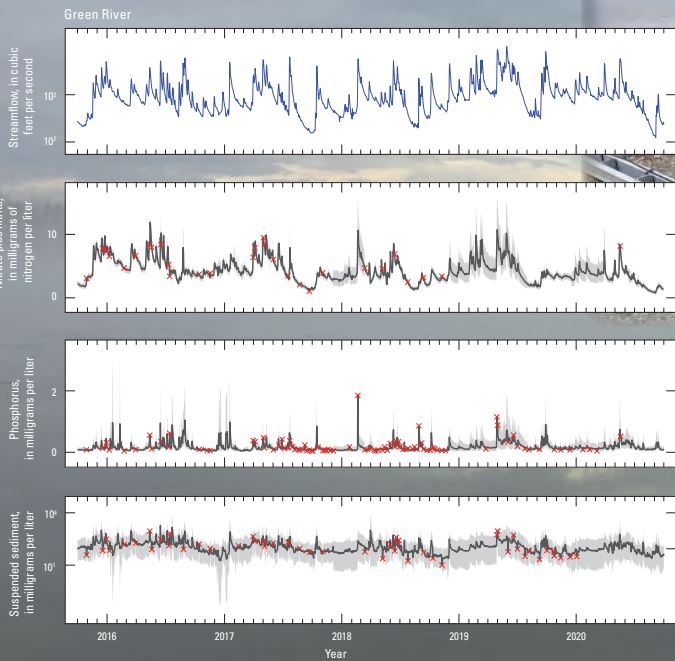


Prepared in cooperation with the Illinois Environmental Protection Agency

Continuous Monitoring and Bayesian Estimation of Nutrient and Sediment Loads from Illinois Watersheds, for Water Years 2016–2020



EXPLANATION
 ■ 95-percent credibility intervals
 — Expected (mean) value
 — Expected (mean) streamflow at gaging station
 × Verification samples

Scientific Investigations Report 2021–5092

Cover:

Background, Photograph of the Illinois River at Florence, Illinois, by David J. Fazio, U.S. Geological Survey

Right, Photograph of the continuous monitoring station on the Kaskaskia River at New Athens, Illinois, by Colin S. Peake, U.S. Geological Survey

Left, Figure 5 of this report

Continuous Monitoring and Bayesian Estimation of Nutrient and Sediment Loads from Illinois Watersheds, for Water Years 2016–2020

By Timothy O. Hodson, Paul J. Terrio, Colin S. Peake, and David J. Fazio

Prepared in cooperation with the Illinois Environmental Protection Agency

Scientific Investigations Report 2021–5092

U.S. Department of the Interior
U.S. Geological Survey

U.S. Geological Survey, Reston, Virginia: 2021

For more information on the USGS—the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment—visit <https://www.usgs.gov> or call 1–888–ASK–USGS.

For an overview of USGS information products, including maps, imagery, and publications, visit <https://store.usgs.gov/>.

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Although this information product, for the most part, is in the public domain, it also may contain copyrighted materials as noted in the text. Permission to reproduce copyrighted items must be secured from the copyright owner.

Suggested citation:

Hodson, T.O., Terrio, P.J., Peake, C.S., and Fazio, D.J., 2021, Continuous monitoring and Bayesian estimation of nutrient and sediment loads from Illinois watersheds, for water years 2016–2020: U.S. Geological Survey Scientific Investigations Report 2021–5092, 40 p., <https://doi.org/10.3133/sir20215092>.

Associated data for this publication:

Hodson, T.O., Terrio, P.J., Peake, C.S., and Fazio, D.J., 2021, Modeled nutrient and sediment concentrations from major rivers in Illinois based on continuous monitoring from October 1, 2015, through September 30, 2020: U.S. Geological Survey data release, <https://doi.org/10.5066/P9MS80IH>.

ISSN 2328-0328 (online)

Acknowledgments

The authors would like to acknowledge the Illinois Environmental Protection Agency and all other entities and individuals who participate on the Illinois Nutrient Loss Reduction Strategy Policy Working Group for their support of this effort. Funding for model development was also provided by the U.S. Geological Survey Uncertainty in Time-Series Observations project of the National Hydrologic Monitoring Program.

Contents

| | |
|--|-----|
| Acknowledgments | iii |
| Abstract | 1 |
| Introduction..... | 1 |
| Purpose and Scope | 2 |
| Methods..... | 2 |
| Data Collection | 4 |
| Covariate-Based Bayesian Imputation | 5 |
| Notation | 5 |
| Model Overview | 6 |
| Observational Model | 6 |
| Process Model | 6 |
| Parameter Models | 7 |
| The Full Joint Distribution Model | 8 |
| Nonidentifiability | 9 |
| Concentration and Load Estimation..... | 9 |
| Limitations | 9 |
| Data Coverage..... | 10 |
| Streamflow and Discrete Water-Quality Data | 10 |
| Imputation Results | 11 |
| Comparison Among Model Forms..... | 13 |
| Loads and Yields | 14 |
| Nitrate | 14 |
| Phosphorus | 15 |
| Suspended Sediment | 15 |
| Continuous Monitoring and Discrete Sampling..... | 15 |
| Network Improvements | 20 |
| Summary..... | 20 |
| References Cited..... | 20 |
| Appendix 1. Station Descriptions | 23 |

Figures

| | |
|---|----|
| 1. Map showing drainage basins and locations of the eight nutrient monitoring stations | 3 |
| 2. Graphs showing mean annual nitrate load at Green River near Geneseo based on <i>A</i> , discrete data with C5-like model <i>B</i> , continuous data with C5-like gap filling and <i>C</i> , continuous data with Bayesian imputation..... | 13 |
| 3. Graphs showing mean annual phosphorus load at the Illinois River near Florence based on <i>A</i> , discrete data with C5-like model <i>B</i> , continuous data with simple surrogate model and <i>C</i> , continuous data with two-fraction model..... | 14 |
| 4. Graphs showing mean annual yields in metric tons per square kilometer for water years 2016–20 for <i>A</i> , nitrate; <i>B</i> , phosphorus; and <i>C</i> , suspended sediment..... | 16 |

| | |
|--|----|
| 5. Graph showing mean annual nitrate load by river based on continuous monitoring with Bayesian imputation (blue) and discrete sampling with C5-like model (red)..... | 16 |
| 6. Graph showing mean annual phosphorus load by river based on continuous monitoring with Bayesian imputation (blue) and discrete sampling with C5-like model (red)..... | 18 |
| 7. Graph showing mean annual suspended-sediment load by river based on continuous monitoring with Bayesian imputation (blue) and discrete sampling with C5-like model (red)..... | 19 |

Tables

| | |
|---|----|
| 1. Continuous nutrient monitoring station information..... | 2 |
| 2. Manufacturer specifications for each instrument..... | 4 |
| 3. Percentage of record with continuous data coverage from October 1, 2015, to September 30, 2020..... | 10 |
| 4. Summary of streamflow during the 2016–20 water years and streamflow relative to the 1980–96 baseline period..... | 11 |
| 5. Analytical method information for discrete water-quality samples..... | 11 |
| 6. Summary of discrete water-quality samples collected during the 2016–20 water years..... | 12 |
| 7. Annual nitrate loads by river estimate based on continuous monitoring with full joint distribution model with Bayesian imputation..... | 15 |
| 8. Annual phosphorus load by river; estimate based on continuous monitoring with Bayesian imputation..... | 17 |
| 9. Annual suspended-sediment load by river; estimate based on continuous monitoring with Bayesian imputation..... | 18 |

Conversion Factors

U.S. customary units to International System of Units

| Multiply | By | To obtain |
|-----------------|-----------|------------------|
| | Length | |
| inch (in.) | 2.54 | centimeter (cm) |

International System of Units to U.S. customary units

| Multiply | By | To obtain |
|-------------------------------------|-----------|--------------------------------|
| | Area | |
| square kilometer (km ²) | 247.1 | acre |
| square kilometer (km ²) | 0.3861 | square mile (mi ²) |
| | Volume | |
| liter (L) | 1.057 | quart (qt) |
| liter (L) | 0.2642 | gallon (gal) |
| cubic meter (m ³) | 264.2 | gallon (gal) |

| Multiply | By | To obtain |
|--|-----------|-------------------------------|
| cubic meter (m ³) | 35.31 | cubic foot (ft ³) |
| Flow rate | | |
| cubic meter per second (m ³ /s) | 70.07 | acre-foot per day (acre-ft/d) |
| Mass | | |
| metric ton (t) | 1.102 | ton, short [2,000 lb] |
| metric ton (t) | 0.9842 | ton, long [2,240 lb] |

Temperature in degrees Celsius (°C) may be converted to degrees Fahrenheit (°F) as follows: °F = (1.8 × °C) + 32.

Datum

Horizontal coordinate information is referenced to the North American Datum of 1983 (NAD 83).

Altitude, as used in this report, refers to distance above the vertical datum.

Supplemental Information

Specific conductance is given in microsiemens per centimeter at 25 degrees Celsius (µS/cm at 25 °C).

Concentrations of chemical constituents in water are given in either milligrams per liter (mg/L) or micrograms per liter (µg/L).

A water year is the period from October 1 to September 30 and is designated by the year in which it ends; for example, water year 2016 was from October 1, 2015, to September 30, 2016.

Abbreviations

| | |
|-----------------------|--|
| Cycle-PO ₄ | WET Labs Cycle orthophosphate |
| FNU | Formazin Nephelometric Unit |
| IEPA | Illinois Environmental Protection Agency |
| MAR | missing at random |
| MCMC | Markov Chain Monte Carlo |
| NLRS | Nutrient Loss Reduction Strategy |
| NO23 | nitrite plus nitrate as nitrogen |
| PO4 | orthophosphate |
| SSC | suspended-sediment concentration |
| TP | total phosphorus |
| USGS | U.S. Geological Survey |
| WY | water year |

Continuous Monitoring and Bayesian Estimation of Nutrient and Sediment Loads from Illinois Watersheds, for Water Years 2016–2020

By Timothy O. Hodson, Paul J. Terrio, Colin S. Peake, and David J. Fazio

Abstract

The State of Illinois is one of the leading contributors of nitrogen, phosphorus, and suspended sediment to the Mississippi River and the Gulf of Mexico. During water years 2016–20, the U.S. Geological Survey, in cooperation with the Illinois Environmental Protection Agency, operated continuous monitoring stations on eight major rivers in Illinois to better quantify nutrient and sediment loadings from the State of Illinois to the Mississippi River. This report estimates nitrate, phosphorus, and suspended-sediment loadings over that period, which can provide a benchmark against which to assess future changes in loading.

In addition, this report develops a new method for incorporating the uncertainty created by gaps in continuous datasets based on Bayesian machine learning. Data gaps are a common problem in continuous monitoring, and gap filling is necessary to quantify loadings and the uncertainty in loadings, which is essential if these results are to provide a benchmark for future studies. The uncertainty estimates may also be useful in an operational context, and this report provides examples of how uncertainty can be used in monitoring-network design and potentially reducing monitoring costs.

Introduction

Drainage basins within the State of Illinois are diverse: from the densely populated Chicago metropolitan area in northeast Illinois to the sparsely populated row-crop agricultural land throughout the State to the Shawnee National Forest in southern Illinois. The various areas with different land uses, particularly the intensely urban and agricultural areas, can contribute substantial inputs of nitrogen, phosphorus, and sediment to receiving streams and rivers. Excess nutrients, a condition known as eutrophication, can promote growth of algae and aquatic vegetation. Then as the algae and aquatic vegetation die and decompose, the water bodies they inhabited may become depleted of oxygen, a condition known as hypoxia. Hypoxia harms or kills fish and other

fauna, affects recreational and commercial use of the water body, and contributes to taste and odor issues in drinking water. Nutrients from Illinois watersheds transported to and down the Mississippi River also contribute to hypoxia in the Gulf of Mexico (Alexander and others, 2008; Heimann and others, 2011; Sprague and others, 2011). In addition to these downstream effects, the loss of nitrogen and phosphorus from agricultural land is costly to farmers and producers, who must amend the loss by applying more fertilizer. Excessive sediment concentrations and loads also contribute to a variety of impairments, such as physical damage to the environment by limiting light penetration and decreasing euphotic zone depth, clogging fish gills, filling in coarse-substrate spawning and feeding areas, smothering invertebrate communities, and potentially consuming available oxygen if the sediment contains organic material. Sediment deposition also causes costly damage to critical infrastructure like barge lanes and reservoirs used for water supply and flood control. Because phosphorus readily sorbs and desorbs to sediment and particulate matter, sediment transport can also transport and release phosphorus, which contributes to eutrophication in many water bodies (Duda, 1985).

Multiple studies have identified Illinois watersheds as having some of the highest yields of nitrogen, phosphorus, and sediment in the Mississippi River Basin (David and Gentry, 2000; Alexander and others, 2008; David and others, 2010; Heimann and others, 2011; Jacobson and others, 2011; Sprague and others, 2011; Robertson and others, 2014; Robertson and Saad, 2019). In response to these high yields, the Illinois Environmental Protection Agency; Illinois Department of Agriculture; and representatives from various governmental, environmental, utility, and stakeholder groups developed the Illinois Nutrient Loss Reduction Strategy (NLRs) in 2015 with the primary goal of improving the management of surface water bodies throughout Illinois and completing the reduction of nitrogen, phosphorus, and sediment delivered to surface water bodies throughout Illinois (Illinois Environmental Protection Agency and others, 2015). The Illinois NLRs was designed as part of a coordinated effort with 10 other states in the Mississippi River Basin to reduce the nutrient load entering the Gulf of Mexico. These efforts

2 Continuous Monitoring and Bayesian Estimation of Nutrient and Sediment Loads from Illinois Watersheds

are expected to benefit local and regional communities and the Nation by helping to reduce losses of nutrients and sediments from the landscape.

As part of the NLRS, the U.S. Geological Survey (USGS), in fiscal and operational cooperation with the Illinois Environmental Protection Agency, began a monitoring program in 2015 to measure and calculate nutrient (nitrate and total phosphorus [TP]) and suspended-sediment loads at monitoring stations on the eight primary Illinois rivers in the Upper Mississippi River Basin. To provide improved estimates of nutrient and sediment loadings, continuous (15-minute frequency) sensors were installed to measure nitrate plus nitrite as nitrogen (NO₃) and orthophosphate (PO₄) concentrations, turbidity, and physicochemical properties (water temperature, pH, specific conductance, and dissolved oxygen). The continuous data record can be used to provide a more complete and accurate record of constituent concentrations and loadings than periodic discrete sampling, particularly during high-streamflow events where conditions can change rapidly.

Purpose and Scope

This report describes 5 years of continuous NO₃, total phosphorus, and suspended-sediment data collected from October 1, 2015, through September 30, 2020, or water years (WY) 2016–20 (hereafter referred to as the study period). This 5-year dataset and accompanying analysis will assess the concentrations, loadings, and characteristics of nutrients and suspended sediment, as well as changes in constituent loads over time and in response to the implementation of

nutrient-reduction land-use practices, wastewater and industrial treatments, and regulatory efforts by comparing our results with the State’s initial baseline nutrient loadings calculated for 1980–96 (hereafter referred to as the baseline). These baseline loadings were established in the Nutrient Reduction Strategy’s Science Assessment (Illinois Environmental Protection Agency and others, 2015). The baseline loadings will be used by NLRS partners to determine progress toward achieving the goal of a 45-percent reduction in nutrient losses across the State of Illinois.

Methods

Eight major river basins in Illinois drain into the Mississippi River (fig. 1). For each of these eight basins, the most downstream USGS gaging station that was not substantially affected by backwater from the Mississippi, Ohio, or Wabash Rivers was chosen. The cumulative drainage areas of these monitoring stations comprise approximately 73.9 percent of the total land area of Illinois (table 1). Parts of some of these basins fall outside of Illinois, and contributions from those areas are included in the load estimates herein: part of the Rock River watershed in Wisconsin, the upstream part of the Kankakee River watershed in Indiana, as well as areas of other smaller watersheds throughout the State. Each site was instrumented with sensors to measure NO₃ and PO₄ concentrations, turbidity, water temperature, pH, specific conductance, and dissolved oxygen on a 15-minute (hereafter referred to as “continuous”) interval.

Table 1. Continuous nutrient monitoring station information.

[USGS, U.S. Geological Survey; ID, identification; km², square kilometer]

| River | USGS ID | Station drainage area (km ²) | Station drainage area in Illinois (km ²) | Basin drainage area in Illinois (km ²) | Percent of station drainage area in Illinois | Areal percent of Illinois |
|----------------------------|----------|--|--|--|--|---------------------------|
| Vermilion | 03339000 | 3,341 | 3,105 | 3,372 | 93 | 2.1 |
| Embarras | 03346500 | 6,042 | 6,042 | 6,307 | 100 | 4.2 |
| Little Wabash ¹ | 03381495 | 8,034 | 8,034 | 8,298 | 100 | 5.5 |
| Rock | 05446500 | 24,732 | 10,290 | 13,789 | 42 | 7.1 |
| Green ² | 05447500 | 2,598 | 2,598 | 2,927 | 100 | 1.8 |
| Illinois ³ | 05586300 | 69,264 | 58,666 | 64,009 | 85 | 40.2 |
| Kaskaskia | 05595000 | 13,439 | 13,439 | 15,045 | 100 | 9.2 |
| Big Muddy | 05599490 | 5,592 | 5,592 | 6,180 | 100 | 3.8 |

¹Drainage area numbers are for the nearby streamgage at Little Wabash River at Carmi, Illinois (U.S. Geological Survey identification number 03381500).

²The Green River is part of the Rock River drainage basin and is included in the 13,789 square kilometer Rock River drainage area in Illinois.

³Drainage area numbers are for the nearby streamgage at Illinois River at Valley City, Illinois (U.S. Geological Survey identification number 05586100).

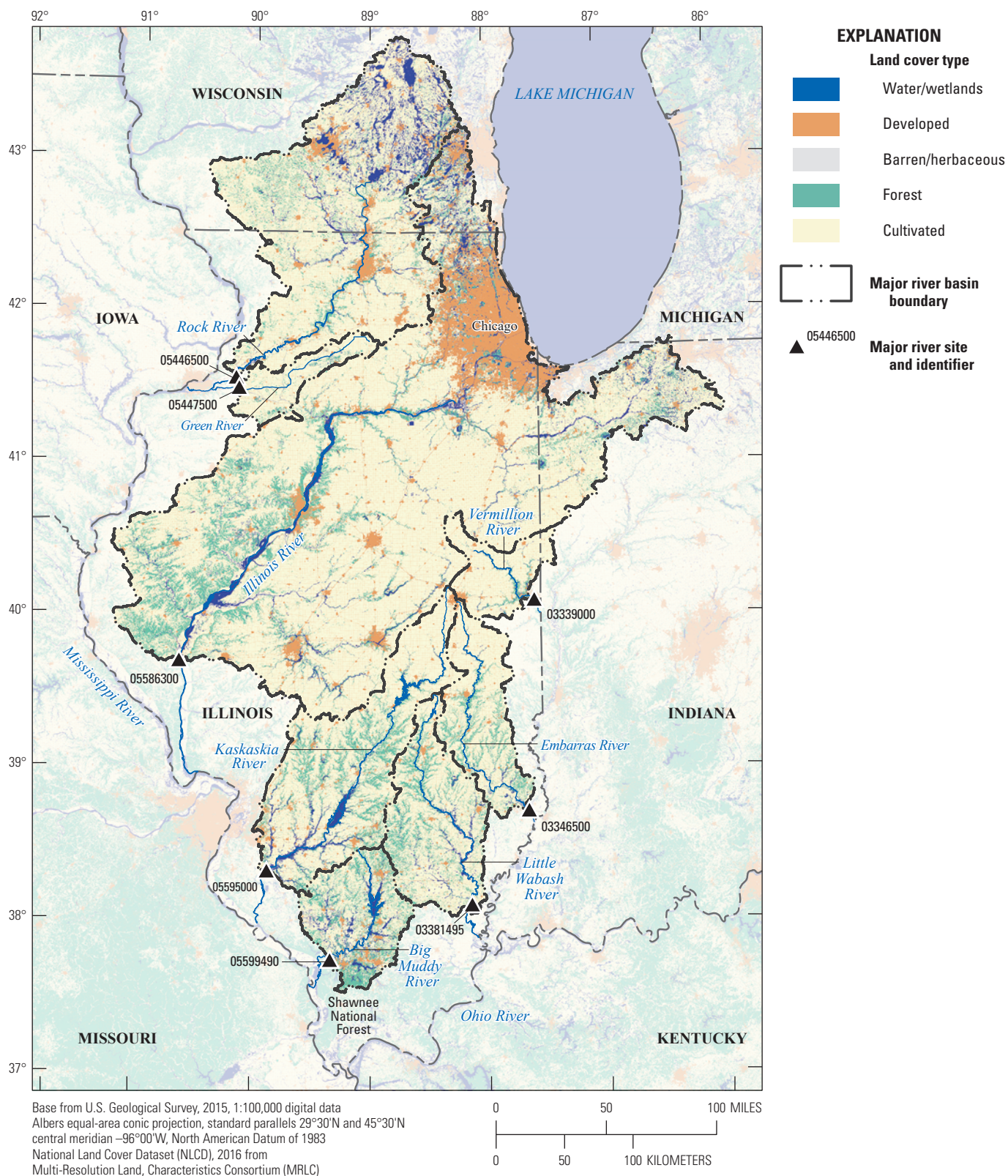


Figure 1. Map showing drainage basins and locations of the eight nutrient monitoring stations.

Data Collection

To the extent practicable, the monitoring instrumentations were in areas of the channel with well-mixed and continuous streamflow, adequate depth under all streamflow conditions, and protection from in-channel debris. Infrastructure at each monitoring station differed depending upon channel and streamflow characteristics, property ownership, and physical support and security requirements. Four of the stations were mounted, at least initially, to the downstream side of existing bridge piers. Two stations were mounted to cylindrical pier or bridge protection cells, and two stations were installed on the side of the channel because of a lack of an adequate bridge pier or other midchannel structure. Initially, the sensors and analyzers were installed in four 4-to-6-inch polyvinyl chloride (PVC) conduits with one conduit each for the YSI multimeter, Hach Nitratex sensor, WET Labs Cycle orthophosphate (Cycle-PO₄) analyzer, and Hach Solitax turbidity sensor (table 2). These configurations changed during the 5-year monitoring period as more advantageous or reliable configurations were determined. These changes are documented in the individual station summaries in appendix 1.

Each group of sensors was at or near an existing USGS streamflow-monitoring station, which provided continuous records of stream stage and flow. Continuous data and annual summaries of the minimum, maximum, and mean concentrations for the water-quality parameters at each monitoring station can be found at the National Water Information System:

Web Interface (https://waterdata.usgs.gov/nwis/wys_rpt/?site_no=03339000, where 03339000 represents the 8-digit USGS station number).

Installations were on the downstream side of bridge supports or pier protection cells at 5 of the 8 sites. Installations failed in some places because of excessive debris and sedimentation. Sites that required specialized infrastructure were the Little Wabash River at Main Street at Carmi, Illinois (03381500), Rock River near Joslin, Ill. (05446500), and Vermilion River near Danville, Ill. (03339000). At Little Wabash River at Carmi, the pipes had to be modified to handle the heavy sediment loads. The Rock River at Joslin site had seasonal problems because of river ice in winter and heavy biological growth in summer. Hanging pipes installed in 2019 helped to mitigate ice damage and biological fouling at this site. Because of abundant debris and rapid sedimentation at the Vermilion River site, the instrumentation had to be relocated away from the bank. In 2017, a gage house was installed on the levee atop the bank with a pumping system to route water up from the river to the instruments inside the gage house. The expansion of the gage house resulted in this station becoming a demonstration site to evaluate the ability of continuous phosphorus analyzers to monitor natural water bodies.

The continuous data record for each parameter was analyzed, approved, and audited according to established USGS protocols and methods (Wagner and others, 2006). These procedures included removal of obvious erroneous data, correction for sensor fouling and calibration drift, and comparison of continuous sensor data with periodic discrete verification

Table 2. Manufacturer specifications for each instrument.

[°C, degree Celsius; ±, plus or minus; mS/cm, millisiemens per centimeter; %, percent; mg/L, milligram per liter; CT, calibration temperature; FNU, Formazin Nephelometric Unit; <, less than; mm, millimeter; mg N/L, milligram of nitrogen per liter; mg P/L, milligrams of phosphorus per liter]

| Parameter | Instrument | Range | Accuracy | Resolution |
|----------------------|--------------------------------|--|--|--|
| Water temperature | YSI EXO2 multiparameter sonde | -5–50°C | -5–35°C: ±0.01°C; 35–50°C: ±0.05°C | 0.001°C |
| Specific conductance | YSI EXO2 multiparameter sonde | 0–200 mS/cm | 0–100 mS/cm: ±0.5%; 100–200 mS/cm: ±1% | 0.0001–0.01 mS/cm, range dependent |
| Dissolved oxygen | YSI EXO2 multiparameter sonde | 0–50 mg/L | 0–20 mg/L: ±1.0%; 20–50 mg/L: ±5.0% | 0.1 mg/L |
| pH | YSI EXO2 multiparameter sonde | 0–14 units | ±0.1 units within ±10°C of CT, ±0.2 units otherwise | 0.01 units |
| Turbidity | YSI EXO2 multiparameter sonde | 0–4000 FNU | 0–999 FNU: ±2%; 1000–4000 FNU: ±5% | 0–999 FNU: 0.01 FNU; 1000–4000 FNU: 0.1 FNU |
| Turbidity | Hach Solitax | 0–4000 FNU | 0–1000 FNU: <1% with calibration | 0.01 FNU |
| Nitrate plus nitrite | Hach Nitratex plus sc 2 mm | 0–50 mg N/L | ±3% or 0.5 mg N/L, whichever is greater | 0.1 mg N/L |
| Orthophosphate | WET Labs Cycle-PO ₄ | 0–0.3 mg P/L; 0.3–1.2 mg P/L possible | 0.0077 mg P/L | 0.001 mg P/L |
| Orthophosphate | YSI P700 | 0.05–15 mg P/L | ±2% or 0.05 mg P/L, whichever is greater | 0.01 mg P/L |

samples that were analyzed by the USGS National Water Quality Laboratory. Mean annual nutrient and suspended-sediment loads were estimated from continuous data using methods described in the next section.

Covariate-Based Bayesian Imputation

Collection of continuous water data has expanded rapidly during the previous decade. With this new style of environmental monitoring come new challenges, one of which is how to quantify the uncertainty in the continuous records. All measurements are imperfect and inherently include some uncertainty. But uncertainty arises from other sources as well, such as from sampling procedures and models used to fill data gaps. Addressing each component of uncertainty in an ad hoc manner can quickly become overwhelming. This problem becomes more manageable, however, using Bayesian methods. Several definitions exist as to what distinguishes a method as “Bayesian,” but most simply, Bayesian methods use Bayes’ theorem to learn about the world from observations and prior knowledge. A key strength of the Bayesian perspective is that each source of uncertainty can be treated as a single phenomenon, such as missing data or data that are partially unobserved. By treating all missing data as random variables, Bayes’ theorem provides a logical means through which to determine the probable values of missing data if given a set of observations and prior knowledge.

Every measurement and, by extension, every inference incorporates some uncertainty arising from missing data. If that uncertainty is not adequately quantified, the measurement is of little practical use. Methods for quantifying uncertainty have evolved in an ad hoc manner among scientific disciplines, and at the time of writing, no well-established “best” practices for quantifying uncertainty in continuous water-quality data exist. One basic approach estimates missing data based on a surrogate. But surrogate data can have gaps as well, leading to studies that use multiple surrogate models, switching among them depending on data availability (Robertson and others, 2018; Lathrop and others, 2019). For example, a study may use turbidity as a surrogate for suspended-sediment concentration (SSC), then use another surrogate, like streamflow, to predict SSC when turbidity data are unavailable in what is essentially a form of analysis (Little and Rubin, 2019).

The exact implementation varies, but typically, a set of potential surrogate models are ranked by goodness of fit. Then, predictions are made by applying the models in rank order: the highest ranked model is used first, then the second ranked model fills in gaps where the first surrogate was missing, and so forth. This multimodel approach can work well for small data gaps, but its shortcomings may become more apparent as gaps increase. Firstly, it is only statistically valid if the data are missing completely at random; in other words, gaps are not dependent on the values of the data (Little and Rubin, 2019). Because the missing completely at random condition is usually unrealistic in practice, predictions made

by combining multiple surrogate models may be biased. While this bias shrinks as the data become more complete, it is impossible to assess the magnitude of the bias. Secondly, the multimodel approach wastes information by not sharing information among the different surrogate models. For example, by regressing SSC on streamflow, the model typically ignores all the information that turbidity data provide about the covariance of SSC and streamflow.

This study develops an alternative approach that uses Bayesian modeling to impute missing data using covariates (in statistics, imputation is the process of replacing missing values with estimates). Bayesian imputation is more resilient against missing data bias; more efficient in terms of information use; and, ultimately, more flexible because of the ease with which it incorporates other sources of uncertainty. The details of the method are described in the following sections, but these are not necessary to interpret the results. Readers interested only in results need only know that the method predicts concentrations of a constituent in a river on the basis of correlations among the target constituent concentration and other surrogates measured during the study, such as water temperature, pH, dissolved oxygen, specific conductance, turbidity, streamflow, season, and time. A more technical summary of the approach is that the model assumes the joint distribution of the data is multivariate lognormal, learns the joint distribution from observations, and then simulates any missing observations using the learned joint distribution (correlations).

Notation

Some notation is helpful to describe the predictive model. Boldface lowercase letters indicate vectors (for example, \mathbf{x}), and lightface lowercase letters represent scalars (σ , z). Boldface uppercase letters indicate matrices (\mathbf{A} , $\mathbf{\Sigma}$). The notation $[a]$ represents the probability of the random variable a , and $[a|b,c]$ represents the conditional probability of a given b and c . This bracket notation provides a compact way to specify stochastic models, but sometimes the distribution needs to be explicit. For example, the following are all equivalent ways of expressing a normally distributed variable x with mean μ and variance σ^2 :

$$[x|\mu, \sigma^2] \quad (1)$$

$$normal(x|\mu, \sigma^2) \quad (2)$$

$$x \sim normal(\mu, \sigma^2) \quad (3)$$

where

- x is a random variable,
- μ is the mean of x , and
- σ^2 is the variance of x .

Deterministic models are denoted by $g()$, with the model arguments contained inside the parentheses, so a simple linear model is represented as

$$g(\boldsymbol{\beta}, x) = \beta_0 + \beta_1 x \quad (4)$$

where

- g is a deterministic model, and
- $\boldsymbol{\beta}$ is a vector containing the model parameters.

Model Overview

This study uses a Bayesian hierarchical model to predict concentrations and loads of water-quality constituents from continuous surrogate observations. The general approach, first described by Berliner (1996), simulates imperfect observational data using three conditionally independent classes of models: process models representing some set of physical processes in the world (these can range from physically based to purely empirical models), observational models representing inaccuracies in measuring the true state of that process, and parameter models representing prior knowledge about the parameters used in the model.

Observational Model

True water quality is never known. At best, we can make an imperfect measurement of the true state. The observational model relates the imperfect measurement to the true state. For example, an optical nitrate sensor measures the concentration of nitrate in a river by the absorbance of ultraviolet light passing through the water. Optical nitrate sensor measurement or any other means of measuring nitrate is subject to observational error. The observational model links the noisy sensor observation (y) and the true concentration (c). Assuming the sensor is affected by a linear bias and noise, the observational model becomes

$$[y|d(\boldsymbol{\beta}, c), \sigma_o^2] \quad (5)$$

$$d(\boldsymbol{\beta}, c) = \beta_o + \beta_1 c \quad (6)$$

where

- y is the observed concentration,
- d models the linear bias,
- c is the true concentration, and
- σ_o^2 is the variance of the noise term.

Some readers may find the following forms, which are other ways of expressing the same model, more familiar:

$$y = \beta_o + \beta_1 c + \epsilon \quad (7)$$

$$\epsilon \sim Normal(0, \sigma_o^2) \quad (8)$$

where

ϵ is the noise term.

To estimate the true concentration (in other words, calibrate the sensor), water samples are collected periodically from the river and analyzed in a lab. These laboratory measurements are assumed to be a perfect measurement of the true concentration in the river. That assumption is, of course, an approximation: laboratory measurements are not perfect, samples get mislabeled, samples degrade between collection and analysis, and the water in the sample may not be representative of the entire depth and width of the river. Nevertheless, if these unobserved sources of uncertainty are small relative to the observed uncertainty, they can be ignored without significantly biasing the estimate of the overall uncertainty. Although not done in this report, each of these sources of uncertainty could be incorporated within a hierarchical framework through additional observation models.

Process Model

The process model describes the factors that affect the true state (z) of the system, in this context, constituent concentrations in a river. The physical and chemical processes affecting the state of a river system are complex. If those processes are known, they can be represented within a process model using physical equations or parameterization. More commonly, those processes are unknown, or at least their boundary conditions are too poorly defined to adequately model them. In such cases, the “process” model is merely an attempt to describe the correlations within a dataset. This latter type of process model is sometimes referred to as descriptive or empirical in other disciplines, but the distinction between empirically and physically based models is poorly defined in practice. In the type of modeling discussed herein, the term process model reflects that the model represents some set of physical processes and is used regardless of the level of physical realism depicted within the model.

Descriptive models tend to generalize poorly, but they can work well for estimating missing data, so long as the assumptions in the model are consistent with the real world. This study adopts the descriptive approach and makes two assumptions in the process model: (1) the concentrations of some constituent in a river (c) reflect a complex interaction of processes, some of which correlate to measurable quantities like streamflow (q); season and time (t); as well as a range of other continuously monitored covariates like water temperature, pH, dissolved oxygen concentration, and turbidity (collectively called \boldsymbol{x}). And (2) the processes not represented

in the model are uncorrelated in time (in other words, the model residuals are uncorrelated in time); although rarely true in time series analyses, this assumption is frequently made to make models more tractable. Under these assumptions, the probability distribution for the true state \mathbf{z} can be written as

$$\mathbf{z} = [\log c, \log q, \sin 2\pi t, \cos 2\pi t, t, \log \mathbf{x}] \quad (9)$$

$$\mathbf{z} \sim \text{lognormal}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad (10)$$

where

- \mathbf{z} is the true state of the system,
- c is concentration of the constituent of interest,
- q is daily mean streamflow,
- t is time in decimal years,
- \mathbf{x} are other covariates,
- $\boldsymbol{\mu}$ is a vector containing the median value for each element in \mathbf{z} , and
- $\boldsymbol{\Sigma}$ is the covariance of \mathbf{z} .

This formulation of the process model has the convenient property that if only c , q , and t are observed (in other words, the only covariate observations are streamflow and time), the model is equivalent to the popular five-parameter model (C5) developed by Cohn and others (1992):

$$\log(c) = \beta_0 + \beta_1 \log(q) + \beta_2 \sin 2\pi t + \beta_3 \cos 2\pi t + \beta_4 t \quad (11)$$

Parameter Models

The parameter models describe the prior distributions of any unobserved components in the model; in essence, they represent knowledge of the world before collecting any observations. To reflect general ignorance of the true model parameters, weakly informative priors were selected. After standardizing the observations, reasonable priors for the process model are

$$\mu_j \sim \text{Normal}(0, 1) \quad (12)$$

$$\boldsymbol{\Sigma} \sim \text{LKJ}(\eta = 2, \sigma_p) \quad (13)$$

$$\sigma_p \sim \text{HalfCauchy}(2.5) \quad (14)$$

where

- μ_j is the prior mean for each of the J covariates in \mathbf{z} ,
- LKJ is a function defining the prior distribution of the covariance matrix $\boldsymbol{\Sigma}$ (Lewandowski and others, 2009),
- η is a hyperprior affecting the shape of the LKJ distribution, $\eta > 1$ is standard, and
- σ_p is a hyperprior on the standard deviation of the LKJ distribution.

The half-Cauchy distribution of σ_p was chosen as a good default, weakly informative prior for hierarchical models (Gelman and others, 2008). The LKJ prior is a prior on the correlation matrix $\boldsymbol{\Sigma}$ and controls the expected amount of correlation among the covariates \mathbf{z} . For $\eta > 1$, the model will favor less correlation, so setting $\eta = 2$ is a means of regularization and helps prevent overfitting in models with several covariates. Weakly informative priors for the observation model are

$$\beta_0 \sim \text{Normal}(0, 1) \quad (15)$$

$$\beta_1 \sim \text{Normal}(1, 1) \quad (16)$$

$$\sigma_o \sim \text{HalfCauchy}(1) \quad (17)$$

where

β_1 has been centered on 1 to reflect that the sensor has been calibrated such that its measurement and true state should have a ratio of approximately 1:1.

The Full Joint Distribution Model

Combining observational, process, and parameter models yields the full joint model, which provides a basis for estimating the posterior distribution (the probability of the unobserved model parameters conditional on the observed data and prior knowledge):

$$\underbrace{[\mu, \sigma_p, \sigma_o, \Sigma, \beta]}_{\text{unobserved}} \underbrace{[x, y]}_{\text{observed}} \underbrace{\propto \prod_i^n \underbrace{[y_i | \beta, \sigma_o, c_i]}_{\text{observation}} \underbrace{[z_i | \mu, \Sigma]}_{\text{process}} \underbrace{[\mu]}_{\text{parameters}} \underbrace{[\Sigma | \sigma_p]}_{\text{parameters}} \underbrace{[\sigma_p]}_{\text{parameters}} \underbrace{[\sigma_o]}_{\text{parameters}}}_{\text{joint}} \quad (18)$$

where

n is the number of observations, and
 i represents the i th observation.

Recalling $\log(c_i) = z_{i,1}$ and lumping c within z and lumping the unobserved parameters into a single variable, θ , the posterior distribution reduces to

$$[\mu, \theta, \Sigma, \beta | x, y] \equiv \underbrace{[\theta | z, y]}_{\text{posterior}} \quad (19)$$

where

θ represents the lumped parameters.

Finding the posterior distribution is only the first step. In this case, learning the model parameters, θ , is only a means to predict the true state of the river (z), which has only been partially observed during the experiment. The next step is to estimate the distribution of the missing data $[z_{mis}, y_{mis}]$ conditional on the observed data $[z_{obs}, y_{obs}]$, also known as the posterior predictive distribution, which is calculated from the posterior by marginalizing over θ :

$$\underbrace{[z_{mis}, y_{mis} | z_{obs}, y_{obs}]}_{\text{posterior predictive}} \propto \int \underbrace{[z_{obs}, y_{obs} | \theta, z_{obs}, y_{obs}]}_{\text{posterior}} \underbrace{[\theta | z_{obs}, y_{obs}]}_{\text{posterior}} d\theta \quad (20)$$

where

mis denotes missing data, and
 obs denotes observed data.

Assuming the missing data are missing at random (MAR), which means the missing data are independent of the observed data, the posterior predictive distribution simplifies to

$$\underbrace{[z_{mis}, y_{mis} | z_{obs}, y_{obs}]}_{\text{posterior predictive}} \propto \int \underbrace{[z_{mis}, y_{mis} | \theta]}_{\text{observ. \times proc.}} \underbrace{[\theta | z_{obs}, y_{obs}]}_{\text{posterior}} d\theta \quad (21)$$

Under the MAR assumption, the probability distribution of the missing data can be estimated by simulating θ from the posterior distribution, then feeding the simulated θ back into the empirical models (forward simulation). Once enough samples are generated, the posterior and posterior predictive distributions are approximately equivalent to the distribution of the samples. If the data are MAR and the Markov Chain Monte Carlo (MCMC) algorithm has converged, these samples will provide

an unbiased approximation of the probability distribution of the missing data. Although in principle MAR is unprovable, in practice, its plausibility is enhanced by including as many covariates as possible to decrease the degree to which gaps depend on unobserved variables (Gelman and others, 2013, p. 450).

Nonidentifiability

The full joint-distribution model described in the previous section was used to predict nitrate concentrations based on measurements made with an optical nitrate sensor (observational model) or other covariates during gaps in the sensor data (process model). A similar type of model was used to predict SSC and TP concentrations but with slight modification; SSC and TP concentrations were predicted using turbidity, as well as other surrogates. Because the relation between the surrogate and the target constituent is nonlinear and potentially heteroskedastic, the observation model uses a lognormal distribution instead of a normal one:

$$s \sim \text{lognormal}(c, \sigma^2) \quad (22)$$

where

s is the surrogate observation.

Then the full joint model becomes

$$\prod_i^n \underbrace{\text{lognormal}([s_i | \boldsymbol{\beta}, \sigma_o, c_i])}_{\text{observation}} \times \underbrace{\text{lognormal}(\mathbf{z}_i | \boldsymbol{\mu}, \boldsymbol{\Sigma}, \sigma_p)}_{\text{process}} \times \underbrace{\dots}_{\text{params.}} \quad (23)$$

Here, processes and observational models are lognormally distributed, so the joint model is no longer identifiable. By omitting the observational model and including the surrogate (s) within \mathbf{x} , the model is made identifiable again. In doing so, the model loses the ability to distinguish between observation uncertainty and process uncertainty, but that sacrifice is unimportant to quantify the overall prediction uncertainty.

Concentration and Load Estimation

The posterior predictive distributions of NO₂, TP, and SSC concentrations were simulated from the full joint model using MCMC with the Python package PyMC3 version 3.8 (Salvatier and others, 2016). A useful property of MCMC is that any quantity derived from a random variable also becomes a random variable; this process is known as equivariance. Because the instantaneous river load can be derived by multiplying concentration by streamflow, the posterior predictive distribution for the load is estimated by simply multiplying the posterior predictive distribution of concentration by a simultaneous streamflow measurement. The uncertainty inherent in those streamflow measurements is ignored because the magnitude of that uncertainty is essentially unknown.

Limitations

The models used in this study make three assumptions that are violated in practice. (1) The continuous data are downsampled to daily means. The observation model assumes the true concentration is also observed as a daily mean, when in fact, the true concentration is only observed as an instantaneous value. The model was initially run without downsampling but that created problems for the MCMC sampler; however, the results were similar to the daily timestep model. (2) The imputation model makes no correction for censoring—that is, samples reported as a less than or greater than value by the laboratory. Censoring in this dataset was minimal, effecting less than 1 percent of the data, so it should not introduce significant bias. Nevertheless, censoring in other studies may be more extensive for other sites or constituents. In those instances, it is up to the investigator to decide whether it is better to correct for censoring bias or missing data bias; at the time of writing, no model fully corrects for both. (3) The model assumes no serial correlation exists among its errors. This simplification is commonly assumed but rarely justified in modeling water-quality timeseries. In general, the bias introduced by this assumption decreases with better surrogates and fewer data gaps. Each of these three limitations could be addressed by building upon the basic modeling approach herein, but that was beyond the scope of this study.

Data Coverage

The percentage of continuous data coverage from October 1, 2015, to September 30, 2020, for each monitoring site is shown in [table 3](#). Data coverages were greater than or equal to 75 percent for 37 out of 64 sensors and analyzers that were deployed at the monitoring sites. Data coverages at Rock and Green Rivers were less than other sites because the sensors were removed during winter to prevent damage by ice flows and because of biological fouling in summer. Little Wabash had the least coverage of all sites because the sensors were either covered with silt during the recession of high-streamflow events or were out of water at low streamflows (less than approximately 24 cubic meters per second).

The Cycle-PO₄ analyzer was used to collect continuous (2-hour interval) PO₄ concentration data. At the start of the project, the Cycle-PO₄ was the only commercially available, battery-powered orthophosphate analyzer. But for the sites in this study, the Cycle-PO₄ proved unreliable and inaccurate at estimating PO₄ concentrations during highly turbid conditions or when PO₄ concentrations exceeded 0.3 milligrams of phosphorus per liter. Cycle-PO₄ analyzers were initially deployed at all eight rivers but were eventually scaled back to just two sites (the Illinois and Kaskaskia Rivers) because of excessive maintenance requirements and manufacturer servicing times. These rivers were prioritized because PO₄ concentrations were strongly correlated with TP concentrations at these rivers, these rivers represent two of the largest watersheds, and these rivers were thought to have the largest phosphorus loads.

In November 2017, a YSI P700 PO₄ analyzer was installed at the Vermilion River station, which was possible because the gage was supplied with alternating current power. Because of its superior filtration system, the YSI P700 was more reliable during turbid conditions, which is why the Vermilion River station had the highest orthophosphate data coverage of any site in the study.

Continuous PO₄ concentration data used in this study were determined using two different onsite analyzers (only one analyzer was used at any site), and the discrete PO₄ concentrations were measured at the USGS National Water Quality Laboratory. Data are available from the National Water Information System (U.S. Geological Survey, 2021). Although analytical procedures differed slightly among instruments, the results are deemed comparable and are hereafter collectively referred to as “orthophosphate” or “PO₄.”

Streamflow and Discrete Water-Quality Data

The water-quality monitoring stations were at established USGS streamgages. Streamflow measurements were made according to established USGS methods and protocols (Turnipseed and Sauer, 2010; Mueller and others, 2013). Computation of the continuous streamflow record, including analysis, approval, and auditing procedures, was also performed according to established USGS methods (Kennedy, 1983).

A summary of streamflow at the eight monitoring sites and a comparison between the 5-year study period (WY 2016–20) and the 1980–96 baseline period is provided in [table 4](#). Streamflows at the monitoring stations were nearly proportional to the upstream drainage areas, as would be expected. Mean streamflow yields ranged from 0.32 to 0.39 cubic meters per square kilometer (m³/km²) at five of the monitoring stations; however, the Rock River (0.47 m³/km²), Embarras River (0.45 m³/km²), and Little Wabash River (0.49 m³/km²) had higher streamflow yields (approximately 25 percent higher). Mean annual streamflows during WY 2016–20 were consistently higher than the mean annual streamflows during the 1980–96 baseline, particularly for the Rock and Green Rivers in the northwest part of the State.

Table 3. Percentage of record with continuous data coverage from October 1, 2015, to September 30, 2020.

[Temp., temperature; SC, specific conductance; DO, dissolved oxygen; NO₂₃, nitrate plus nitrite; PO₄, orthophosphate]

| River | Temp. | SC | DO | pH | Turbidity (YSI) | Turbidity (Hach) | NO ₂₃ | PO ₄ |
|---------------|-------|----|----|----|-----------------|------------------|------------------|-----------------|
| Vermilion | 80 | 80 | 79 | 75 | 79 | 76 | 77 | 178 |
| Embarras | 99 | 99 | 99 | 89 | 99 | 85 | 87 | 9 |
| Little Wabash | 53 | 52 | 41 | 41 | 49 | 61 | 62 | 5 |
| Rock | 69 | 67 | 67 | 64 | 58 | 57 | 68 | 2 |
| Green | 75 | 73 | 65 | 69 | 61 | 58 | 73 | 3 |
| Illinois | 98 | 93 | 98 | 89 | 97 | 90 | 94 | 29 |
| Kaskaskia | 98 | 95 | 96 | 95 | 98 | 94 | 95 | 32 |
| Big Muddy | 91 | 91 | 91 | 82 | 86 | 84 | 78 | 25 |

¹Orthophosphate analyzer upgraded to YSI P700 on November 1, 2017.

Table 4. Summary of streamflow during the 2016–20 water years and streamflow relative to the 1980–96 baseline period.

| River | 2016 | 2017 | 2018 | 2019 | 2020 | 2016–20 mean | 1980–96 mean | Change from 1980–96 to 2016–20, in percent |
|---------------|--|---------|--------|---------|---------|-----------------|--------------------|---|
| | Annual streamflow, in millions of cubic meters | | | | | | | |
| Vermilion | 1,361 | 1,182 | 1,255 | 1,526 | 1,188 | 1,303 | 1,143 | 14 |
| Embarras | 2,510 | 2,428 | 2,348 | 3,623 | 3,192 | 2,820 | ¹ 1,943 | 45 |
| Little Wabash | 3,848 | 2,501 | 3,211 | 5,503 | 4,580 | 3,929 | 2,856 | 38 |
| Rock | 9,079 | 11,742 | 9,764 | 15,360 | 12,308 | 11,651 | 6,846 | 70 |
| Green | 1,052 | 833 | 758 | 1,618 | 1,002 | 1,052 | 701 | 50 |
| Illinois | 27,392 | 26,085 | 20,449 | 37,553 | 33,238 | 28,944 | 23,437 | 23 |
| Kaskaskia | 5,247 | 4,200 | 3,934 | 7,962 | 7,267 | 5,722 | ¹ 4,086 | 40 |
| Big Muddy | 2,887 | 1,661 | 1,995 | 3,288 | 2,327 | 2,432 | 1,975 | 23 |
| Total | 106,755 | 101,266 | 87,428 | 152,866 | 130,205 | 115,704 | 42,987 | Mean =35 |

¹Estimate based on upstream site and scaled by drainage area ratio.

Discrete water-quality samples were collected routinely at each site during maintenance visits and storm events. Data are available from the National Water Information System (U.S. Geological Survey, 2021). These samples were collected to provide direct measures of nutrients and sediment to develop surrogate models and quantify errors in sensor data. Analytical method information used for the discrete water-quality samples is provided in [table 5](#). Discrete samples were typically collected using a weighted-bottle sampler with a 1-liter precleaned polypropylene bottle suspended by a rope and raised and lowered by hand. The sampler was lowered as quickly as possible to the location and depth of the continuous sensors, where the bottle was allowed to fill, thereby obtaining a water sample from as close to the sensor as possible. Periodically, depth- and width-integrated cross-sectional samples were also collected for nutrients and suspended sediment. These samples were used to characterize whether water passing by the sensor was representative of the cross section of the river. Particular effort was made to collect cross-section samples during high-streamflow events when channel and streamflow conditions can differ substantially from low and normal streamflow periods. A corresponding weighted-bottle

point sample at the sensor location was collected with the cross-section samples for comparison purposes. Relative percent differences between all point and cross-section samples had a mean value of 8.8 and a median value of 6.1, and almost no sites had consistent biases, indicating that the samples were representative of the full river transect. Only SSC at the Illinois River station showed a general bias of higher concentrations in the point samples (median relative percent difference of 10.8). Summary statistics of the discrete sample concentrations at each of the eight nutrient monitoring stations are shown in [table 6](#).

Imputation Results

[Figure 2](#) compares the posterior nitrate loads estimated at the Green River using three approaches. The term “posterior” refers to the fact that the probability distribution was calculated after considering the data; it can be thought of as the uncertainty in the result based on the set of observations and the model. The three approaches are (A) using only

Table 5. Analytical method information for discrete water-quality samples.

[μm , micrometers; mg/L, milligrams per liter; sulfuric acid, H_2SO_4 ; <, less than]

| Parameter | Preservation | Analysis method | Limit detection method |
|----------------------------------|--|---|------------------------|
| Nitrate plus nitrite as nitrogen | Filtered (0.45 μm), chilled, dark bottle | Colorimetry, enzymatic reduction-diazotization (Patton and Kryskalla, 2011) | 0.04 mg/L |
| Total phosphorus | Chilled, H_2SO_4 acid to pH<2 | Colorimetry, alkaline persulfate digestion (Fishman, 1993) | 0.01 mg/L |
| Orthophosphate | Filtered (0.45 μm), chilled, dark bottle | Colorimetry, phosphomolybdate reduction (Patton and Kryskalla, 2003) | 0.004 mg/L |

Table 6. Summary of discrete water-quality samples collected during the 2016–20 water years. Data available from the National Water Information System (U.S. Geological Survey, 2021).

[mg/L, milligrams per liter]

| Constituent | River | Sample count | Minimum (mg/L) | Maximum (mg/L) | Median (mg/L) | 25th percentile (mg/L) | 75th percentile (mg/L) |
|----------------------------------|---------------|--------------|----------------|----------------|---------------|------------------------|------------------------|
| Nitrate | Vermilion | 99 | 0.54 | 10.80 | 5.58 | 3.71 | 6.84 |
| | Embarras | 72 | 0.04 | 6.92 | 1.84 | 0.47 | 3.27 |
| | Little Wabash | 72 | 0.06 | 4.80 | 0.84 | 0.43 | 1.31 |
| | Rock | 32 | 2.11 | 6.61 | 4.45 | 3.57 | 5.55 |
| | Green | 33 | 1.03 | 9.50 | 4.70 | 3.48 | 7.48 |
| | Illinois | 177 | 1.02 | 6.78 | 3.72 | 2.91 | 4.50 |
| | Kaskaskia | 68 | 0.04 | 3.48 | 0.60 | 0.34 | 1.29 |
| | Big Muddy | 70 | 0.05 | 1.28 | 0.39 | 0.27 | 0.58 |
| Total phosphorus | Vermilion | 103 | 0.07 | 1.69 | 0.25 | 0.15 | 0.49 |
| | Embarras | 72 | 0.07 | 1.17 | 0.29 | 0.18 | 0.47 |
| | Little Wabash | 73 | 0.08 | 1.27 | 0.31 | 0.18 | 0.49 |
| | Rock | 94 | 0.06 | 0.65 | 0.19 | 0.14 | 0.24 |
| | Green | 97 | 0.04 | 1.85 | 0.1 | 0.07 | 0.2 |
| | Illinois | 181 | 0.12 | 1.07 | 0.32 | 0.24 | 0.4 |
| | Kaskaskia | 64 | 0.13 | 1 | 0.38 | 0.27 | 0.63 |
| | Big Muddy | 73 | 0.12 | 1.24 | 0.22 | 0.17 | 0.3 |
| Suspended-sediment concentration | Vermilion | 101 | 1 | 2,210 | 121 | 25 | 499 |
| | Embarras | 63 | 14 | 1,240 | 169 | 57 | 443 |
| | Little Wabash | 69 | 14 | 3,510 | 108 | 46 | 306 |
| | Rock | 49 | 9 | 486 | 70 | 48 | 89 |
| | Green | 52 | 11 | 950 | 92 | 62 | 225 |
| | Illinois | 201 | 19 | 1,700 | 86 | 63 | 134 |
| | Kaskaskia | 68 | 16 | 925 | 83 | 44 | 221 |
| | Big Muddy | 70 | 8 | 929 | 71 | 37 | 140 |

the discrete data with the C5-like model discussed in the “Methods” section, which is equivalent to the C5 model with regularization (that is, the addition of constraints to reduce overfitting); (B) using continuous data with the standard method imputation (also discussed in the “Methods” section), which is to fill gaps with the C5-like model; and (C) using the method developed in this study, which fills gaps in the continuous data with Bayesian imputation based on covariates.

At the Green River, continuous monitoring with the standard method of filling gaps (B) offered only modest improvement over discrete sampling. The poor performance of continuous monitoring is in part due to the extent of gaps at this site but also their timing. Like many continuous monitoring sites, data gaps were more frequent during periods of high streamflow because instruments are more prone to damage or foul during high streamflow. For the sites and constituents in this study, most loading occurs during short-duration storm events, and even short data gaps during these critical periods

can contribute substantial uncertainty to load estimates. Because the standard method of imputing gaps presumes that gaps occur at random, which is typically untrue in practice, this method can yield biased results.

Another shortcoming of the standard approach is its reliance on a small number of discrete samples to impute the missing periods, whereas the Bayesian approach learns the full joint distribution, and thereby incorporates all the information from continuous monitoring when imputing data. Because of these two factors, Bayesian imputation provides a more precise and less biased estimate of loading, so long as the MCMC sampling properly converges, which can be assessed through standard metrics (for review, see Gelman and others, 2013).

For the Green River example shown in figure 2, Bayesian imputation (C) yielded significantly lower mean annual loads and lower uncertainty than the standard method of filling gaps in continuous data (B). In general, however, the difference between these methods will depend on the distribution

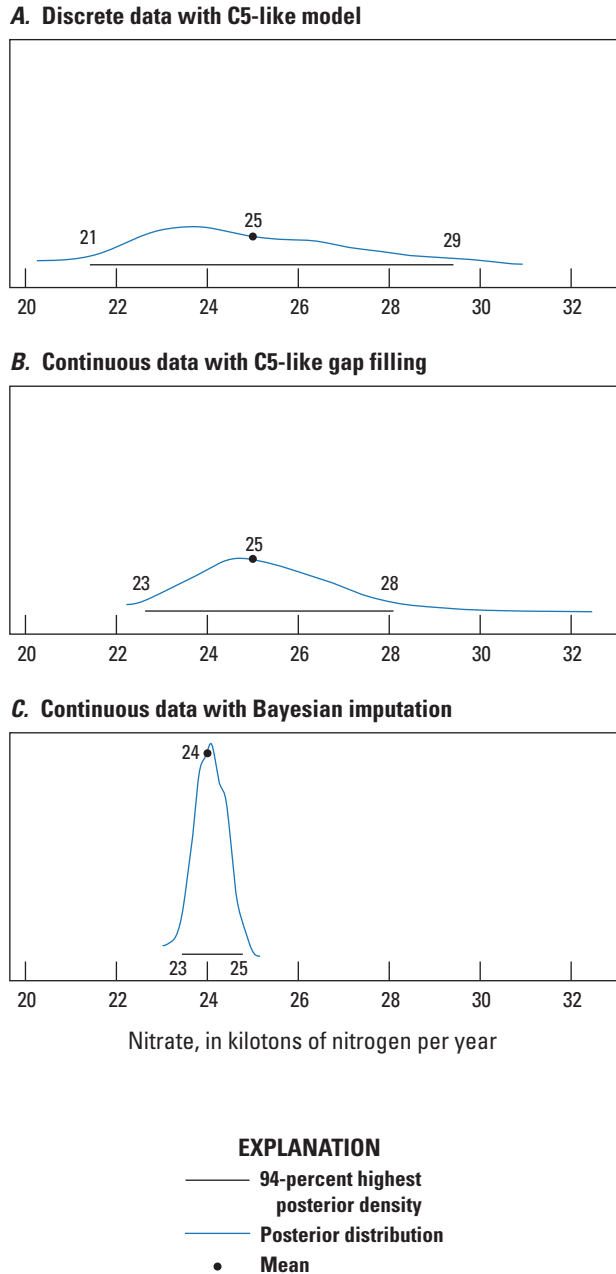


Figure 2. Graphs showing mean annual nitrate load at Green River near Geneseo based on *A*, discrete data with C5-like model *B*, continuous data with C5-like gap filling and *C*, continuous data with Bayesian imputation. The black line shows the 94-percent highest posterior density of that distribution, which is approximately the 94-percent confidence interval in this case, and the blue line represents the posterior predictive distribution obtained by the Markov Chain Monte Carlo algorithm.

and extent of missing data. When data gaps are small, the two methods will produce similar results. If missing data bias causes the standard method to underestimate uncertainty, Bayesian imputation may yield a larger, albeit more accurate, uncertainty estimate.

Comparison Among Model Forms

The nitrate and suspended-sediment models used in this study share a similar structure with one key difference: nitrate was measured by a nitrate sensor, which is designed and calibrated to respond linearly to changes in nitrate concentration. By comparison, none of the sensors directly measure SSC, and turbidity, its best proxy, responds nonlinearly to changes in suspended-sediment concentration. This difference is reflected in the structure of the respective models. In the nitrate model, the linear nitrate sensor is represented in the observational model (eqs. 1–3), whereas the other surrogates are represented in the lognormal process model. Conversely, the SSC model has no observational model, only the surrogates represented in the lognormal process model.

Although the simpler lognormal model could be used for nitrate, it overestimates uncertainty, especially at high concentrations, because it treats the nitrate sensor as a nonlinear surrogate for the true nitrate concentration. Thus, using expert knowledge to structure the model to reflect the data-generating process more accurately can dramatically reduce uncertainty without requiring additional data.

Once onsite TP analyzers become commercially available, an observational model could be used to estimate TP. However, with the technology available during the study, only the dissolved fraction (PO₄) could be measured onsite, whereas the particulate fraction must be estimated using discrete sampling or using a surrogate, like turbidity. Neither of the previous models are ideally suited for this case. The initial approach to modeling TP was essentially a hybrid of the two models: an observational model was used to represent the in situ measurements of the PO₄ fraction and a lognormal surrogate model represented the particulate fraction. But, unlike nitrate, the additional complexity failed to substantially reduce uncertainty in load estimates. Figure 3 compares phosphorus loads from the Illinois River estimated by (A) the discrete data with a C5-like model, (B) continuous data with the simple surrogate model, and (C) continuous data with the two-fraction model. Only the Illinois River, Kaskaskia River, and Vermilion River stations had enough PO₄ data to use the two-fraction model, and in both cases, results from the two-fraction model were similar to those obtained by a simpler model. Moreover, omitting PO₄ data entirely did not substantially affect load estimates at the monitoring stations in this study.

Models were only evaluated for their uncertainty in annual loads. In other contexts, such as predicting instantaneous concentration, the two-fractional model might outperform the simpler model. A previous study at the Illinois River

station showed that including continuous PO₄ data improved predictions of TP concentration (Terrio and others, 2015), so it was thought PO₄ data would reduce load uncertainty as well. The fact that it did not likely reflects the sources and processes affecting phosphorus transport in these watersheds—in general, periods of high PO₄ concentration were associated with below-mean streamflow and did not contribute much load. In addition, the PO₄ analyzers were unreliable during high streamflow when most loading typically occurs, so the most critical periods of record were missing. Different hydrologic conditions, as well as the development of more reliable PO₄ analyzers and, eventually, TP analyzers, will affect model performance in future applications, and further research could explore the factors affecting whether a particular model or set of instruments is likely to perform better than another.

Loads and Yields

The following sections present annual and mean annual loads and yields of nitrate, phosphorus, and sediment for each monitoring station during WY 2016–20 on the basis of results from the full joint-distribution model. Estimated daily loads generated by the model are available in an accompanying data release (Hodson and others, 2021). To assess the benefit of continuous monitoring relative to discrete sampling, loads were also estimated using only discrete sampling data and streamflow with the same model, which is equivalent to the classic C5 model (Cohn and others, 1992) with regularization. Load estimates using continuous and discrete datasets are shown in figures 4–7 in the following sections. Each figure depicts the posterior probability distribution of the mean annual load during the 2016–20 water years. Additional comparison of the models will be discussed in the “Model Comparison” section.

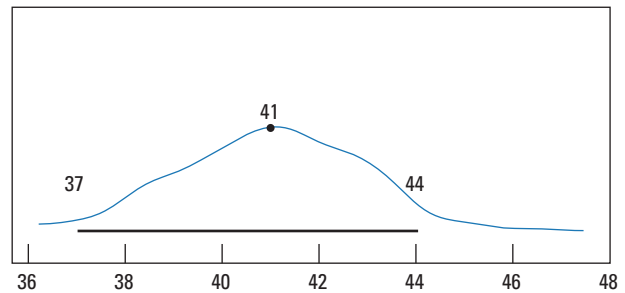
Lastly, results from this study are used to assess how nutrient loads from the eight Illinois rivers have changed relative to those during the 1980–96 baseline period. Loads for the baseline period were estimated by David and others (2015) using discrete sample data and different models, so any comparisons between the baseline and this study may be biased by differences in methodology (see figs. 5–7 and the “Comparison Among Model Forms” section). Correcting such bias is an important topic for future research but was beyond the scope of this study. With that important caveat, the following sections provide a brief overview of some of the largest apparent differences between loadings during the study and baseline periods.

Nitrate

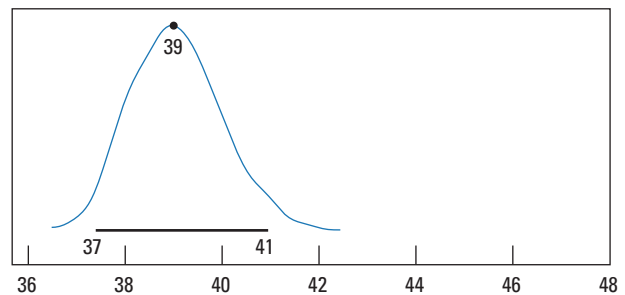
Mean annual nitrate loads and yields for WY 2016–20 ranged from 1.3 to 113.6 metric kilotons and 0.23 to 2.57 metric tons per square kilometer, respectively (table 7, figs. 4 and 5). Most notable is that the Vermilion River had the largest

nitrate yield and fourth largest load but is only 2.5 percent of all the basin areas. Like mean annual streamflow, the mean annual nitrate loads during the 2016–20 period were consistently higher than those during the 1980–96 baseline

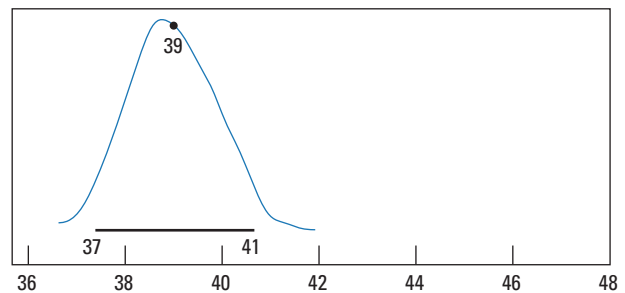
A. Discrete data with C5-like model



B. Continuous data with simple surrogate model



C. Continuous data with two-fraction model



Phosphorous, in kilotons per year

EXPLANATION

- 94-percent highest posterior density
- Posterior distribution
- Mean

Figure 3. Graphs showing mean annual phosphorus load at the Illinois River near Florence based on *A*, discrete data with C5-like model *B*, continuous data with simple surrogate model and *C*, continuous data with two-fraction model. The black line shows the 94-percent highest posterior density of that distribution, which is approximately the 94-percent confidence interval in this case, and the blue line represents the posterior predictive distribution obtained by the Markov Chain Monte Carlo algorithm.

period, except for the Vermilion and Embarras Rivers. The Illinois River had a small 7-percent increase in the mean annual nitrate load compared to a 19-percent increase in mean annual streamflow, indicating that concentration dropped during the same period. The Rock River had the second largest nitrate load and largest increase in nitrate load; however, the 78-percent increase in nitrate load corresponded to a 70-percent increase in mean annual streamflow.

Phosphorus

Mean annual phosphorus loads and yields for WY 2016–20 ranged from 0.3 to 9.7 metric kilotons and 0.10 to 0.19 metric tons per square kilometer, respectively (table 8, figs. 4 and 6). Most notable is that the Kaskaskia River had the largest phosphorus yield and second largest load but is only 10.1 percent of the total basin areas. Like mean annual streamflow, the mean annual phosphorus loads during the 2016–20 period were consistently higher than those during the 1980–96 baseline period, except for the Vermilion and Green Rivers. The Rock River had a small 14-percent increase in the mean annual phosphorus load compared to a 70-percent increase in mean annual streamflow. The Kaskaskia and Big Muddy Rivers had the largest increase in loads (98 and 75 percent, respectively) corresponding to only 40 and 23 percent, respectively, increases in mean annual streamflow.

Suspended Sediment

Mean suspended-sediment loads and yields for WY 2016–20 ranged from 208 to 4,112 kilotons and 33 to 168 metric tons per square kilometer, respectively (table 9, figs. 4 and

7). Most notable is that the Embarras River had the largest suspended-sediment yield and second largest load but is only 4.5 percent of the total basin areas. Also, even though the Kaskaskia River had the highest phosphorus yield, it had the fourth lowest sediment yield, whereas the Rock River had the lowest phosphorus and sediment yield.

Continuous Monitoring and Discrete Sampling

To assess the added value in having continuous monitoring over discrete sampling alone, the imputation model was first trained using the continuous dataset, then the model was retrained using only discrete observations and continuous streamflow, and the two results were compared (note that the models used to generate these results differ from those used by the Illinois NLRS, which were not evaluated in this comparison). In this comparison, the discrete-data model was simple. A more sophisticated model may perform better, but the same argument could be made of the continuous-data model. By using the same model with both datasets, we attempt to make a fair comparison of the monitoring strategies. In this study, the value of continuous monitoring was assessed on mean loads and loading uncertainty, but in practice, monitoring networks serve multiple goals, which could all factor into the value of the monitoring network.

The posterior mean annual loads for the continuous and discrete datasets are shown in figures 5–7. In general, loads were lower, and uncertainties were smaller for the models trained on continuous data. However, individual monitoring stations contributed different amounts of uncertainty to the

Table 7. Annual nitrate loads by river estimate based on continuous monitoring with full joint distribution model with Bayesian imputation.

[km², square kilometer; WY, water year; NA, not applicable]

| River | Annual nitrate load (metric kilotons) | | | | | | 2016–20 mean | 1980–96 mean | Mean annual yield (metric tons per km ²) 2016–20 mean |
|---------------|---------------------------------------|---------|---------|---------|---------|-------|-------------------|--------------|---|
| | 2016 WY | 2017 WY | 2018 WY | 2019 WY | 2020 WY | | | | |
| Vermilion | 8.1 | 8.7 | 10.1 | 8.5 | 7.6 | 8.6 | 9.1 | 2.57 | |
| Embarras | 7.8 | 12.1 | 5.0 | 12.1 | 7.1 | 8.8 | ¹ 10.4 | 1.46 | |
| Little Wabash | 3.5 | 2.5 | 2.5 | 5.4 | 2.2 | 3.2 | 2.8 | 0.40 | |
| Rock | 41.2 | 54.3 | 36.5 | 58.5 | 46.0 | 47.3 | 26.6 | 1.91 | |
| Green | 6.6 | 4.8 | 3.5 | 9.6 | 4.1 | 5.7 | 3.6 | 2.19 | |
| Illinois | 124.0 | 112.6 | 71.6 | 145.6 | 114.2 | 113.6 | 106.6 | 1.64 | |
| Kaskaskia | 5.6 | 3.7 | 5.6 | 8.3 | 7.4 | 6.1 | ² 5.4 | 0.45 | |
| Big Muddy | 1.3 | 0.9 | 1.2 | 1.6 | 1.6 | 1.3 | 0.9 | 0.23 | |
| Total | 198.0 | 199.6 | 136.0 | 249.5 | 190.1 | 194.6 | 165.4 | NA | |

¹Estimate based on upstream station (U.S. Geological Survey identification number 03345500) and scaled by drainage area ratio.

²Estimate based on upstream station (U.S. Geological Survey identification number 05594100) and scaled by drainage area ratio.

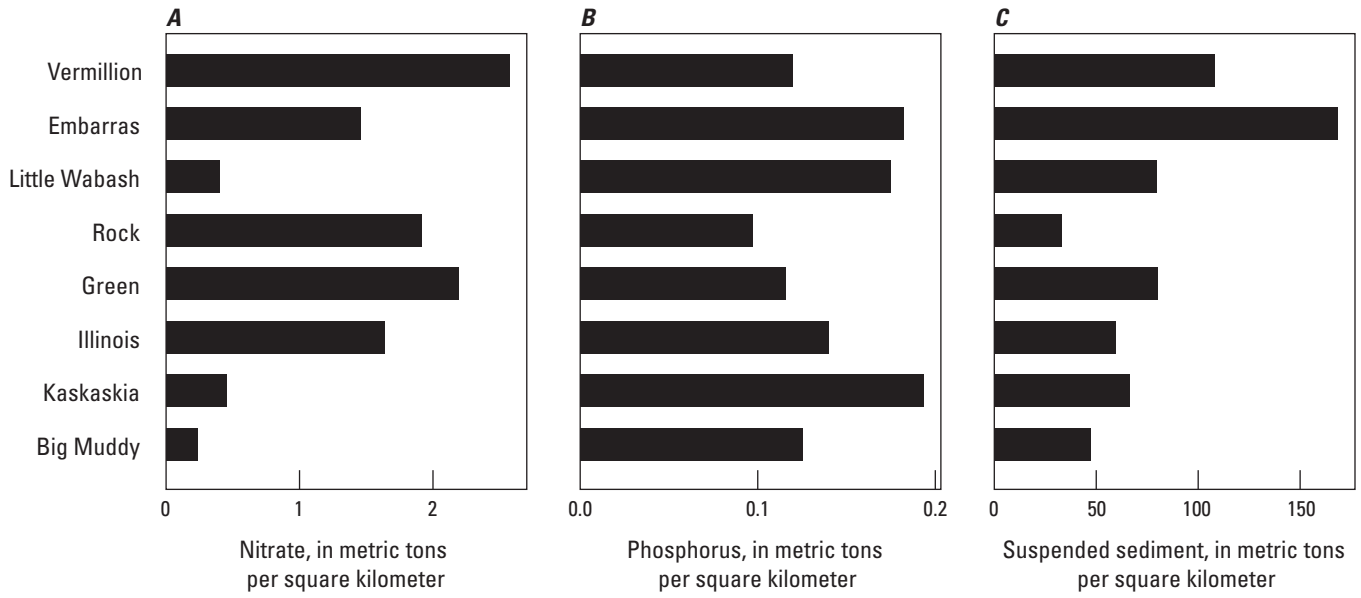


Figure 4. Graphs showing mean annual yields in metric tons per square kilometer for water years 2016–20 for *A*, nitrate; *B*, phosphorus; and *C*, suspended sediment. Estimate is based on continuous monitoring with Bayesian imputation.

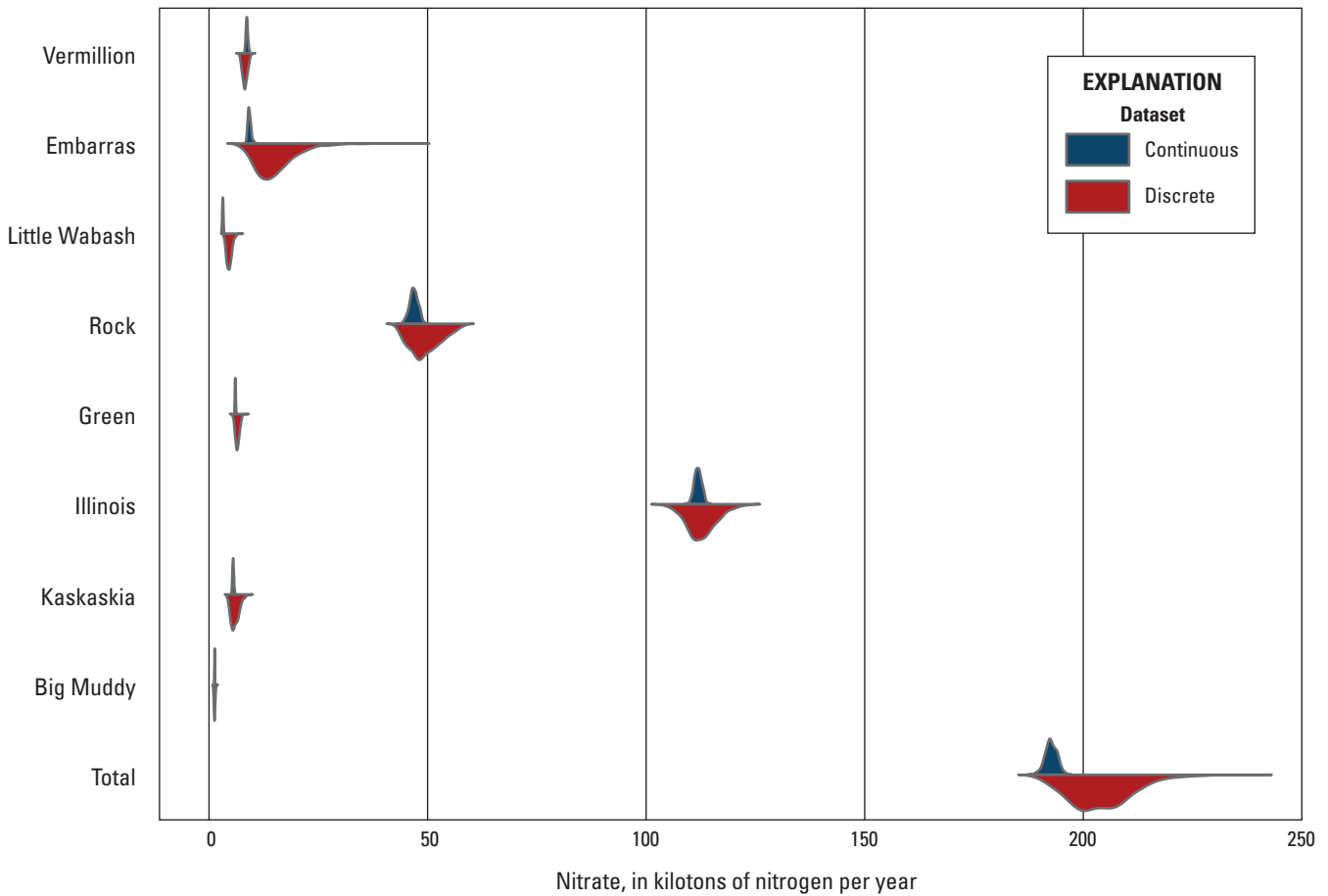


Figure 5. Graph showing mean annual nitrate load by river based on continuous monitoring with Bayesian imputation (blue) and discrete sampling with C5-like model (red).

Table 8. Annual phosphorus load by river; estimate based on continuous monitoring with Bayesian imputation.

[km², square kilometer; WY, water year]

| River | Annual phosphorus load (metric kilotons) | | | | | | | Mean annual yield (metric tons per km ²) 2016–20 mean |
|---------------|--|---------|---------|---------|---------|--------------|--------------|---|
| | 2016 WY | 2017 WY | 2018 WY | 2019 WY | 2020 WY | 2016–20 mean | 1980–96 mean | |
| Vermilion | 0.5 | 0.3 | 0.4 | 0.5 | 0.4 | 0.4 | 0.4 | 0.12 |
| Embarras | 1.1 | 0.9 | 0.8 | 1.5 | 1.4 | 1.1 | 0.8 | 0.18 |
| Little Wabash | 1.3 | 0.8 | 1.0 | 1.9 | 1.9 | 1.4 | 0.9 | 0.18 |
| Rock | 1.8 | 2.6 | 2.1 | 3.1 | 2.4 | 2.4 | 2.1 | 0.10 |
| Green | 0.3 | 0.2 | 0.2 | 0.5 | 0.2 | 0.3 | 0.3 | 0.12 |
| Illinois | 10.2 | 8.7 | 7.6 | 11.7 | 10.2 | 9.7 | 7.4 | 0.14 |
| Kaskaskia | 2.4 | 1.7 | 1.7 | 3.5 | 3.7 | 2.6 | 1.3 | 0.19 |
| Big Muddy | 0.7 | 0.5 | 0.6 | 0.8 | 0.7 | 0.7 | 0.4 | 0.13 |
| Total | 18.3 | 15.6 | 14.4 | 23.6 | 20.8 | 18.5 | 13.7 | NA |

¹Estimate based on upstream station (U.S. Geological Survey identification number 03345500) and scaled by drainage area ratio.

²Estimate based on upstream station (U.S. Geological Survey identification number 05594100) and scaled by drainage area ratio.

overall total. In our comparison, the overall load estimated from the discrete data was biased high relative to the continuous data for every constituent: 7 percent for nitrate, 9 percent for phosphorus, and 30 percent for suspended sediment (figs. 5–7). Whether or not a monitoring network can accurately estimate a quantity is intrinsic to the network’s value. Furthermore, such bias presents an important consideration for continuous monitoring networks, including this effort in Illinois. In the past, water-quality monitoring was primarily carried out by discrete sampling. As continuous monitoring becomes more commonplace, loads could appear to decrease over time because of the shift in monitoring approaches; therefore, comparisons between periods of continuous and discrete data need to be made with care. For that reason, this report only highlights some of the largest apparent loading trends and avoids examining them in detail.

The other benefit of continuous monitoring was that it tended to reduce the uncertainty associated with load estimates. Continuous monitoring reduced uncertainty more at some sites than others. Because reducing uncertainty is equivalent to gaining information, the previous statement also means continuous monitoring added more information for some sites than for others. In figures 5–7, the widths of the posterior distributions are analogous to the uncertainty in each model. Without being mathematically rigorous, the difference in widths between the discrete- versus the continuous-based probability distributions can be thought of as the difference in uncertainty, or the information gain, that resulted from continuous monitoring. If the distribution of the continuous model was narrower than that of the discrete model, then continuous monitoring added information. If means and widths of both distributions were similar, then continuous monitoring potentially added little-to-no additional information.

Streamflow was an important factor affecting the information gain from continuous monitoring. Recall that load is determined by multiplying concentration by streamflow. Therefore, if all sites had similar uncertainty in concentration, that uncertainty is then multiplied by streamflow in determining loading, and as a result, the largest rivers will also tend to have the largest uncertainty in load. Consequently, loads in larger rivers tend to benefit most from continuous monitoring in the sense that these sites add the most information to the overall load estimate. All other things being equal, when planning a continuous monitoring network for monitoring loads, giving the largest watersheds priority may be a beneficial approach. This runs counter to some conventional thinking that assumes that because conditions on large rivers typically change gradually, less frequent monitoring may suffice. Smaller rivers tended to contribute smaller loads, though not always, and as a result, benefited less from continuous monitoring. However, because of their greater variability in concentration and streamflow, smaller rivers may benefit more from continuous monitoring in other contexts. Changes in concentrations and loads resulting from best management practices are generally observed sooner in small watersheds compared to large watersheds, and continuous monitoring can help quantify these changes more rapidly and with less uncertainty. Continuous monitoring can also be advantageous in monitoring for and predicting extremes. Many machine-learning techniques and numerical models rely on some form of regularization during the calibration process. Regularization helps to reduce prediction error by biasing all estimates toward the mean, but this can adversely affect other properties of the distribution, such as the frequency and magnitude of extreme events. Collecting more data, as in continuous monitoring,

18 Continuous Monitoring and Bayesian Estimation of Nutrient and Sediment Loads from Illinois Watersheds

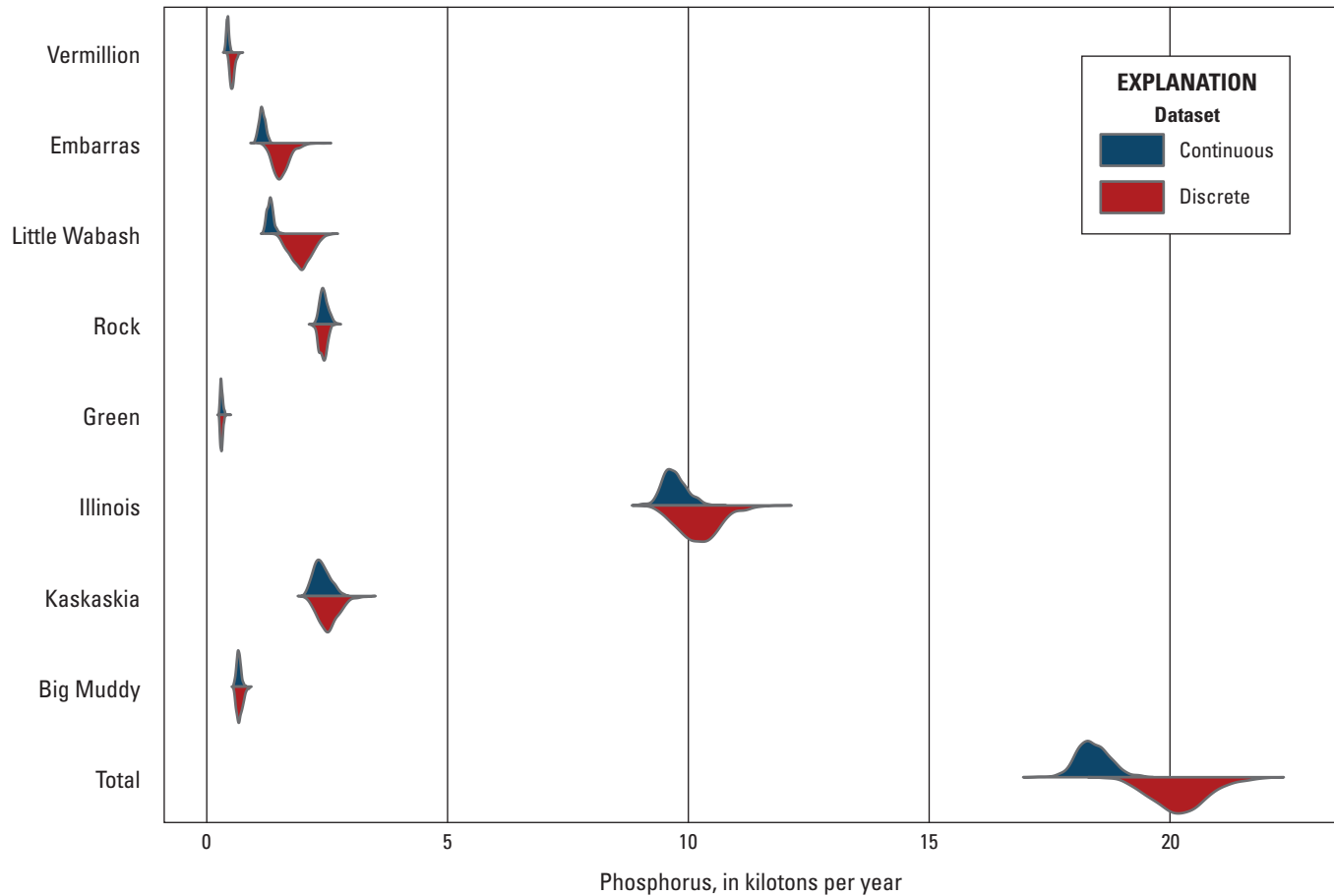


Figure 6. Graph showing mean annual phosphorus load by river based on continuous monitoring with Bayesian imputation (blue) and discrete sampling with C5-like model (red).

Table 9. Annual suspended-sediment load by river; estimate based on continuous monitoring with Bayesian imputation.

[km², square kilometers; WY, water year; NA, not available]

| River | Annual suspended-sediment load (metric kilotons) | | | | | | Mean annual yield (metric tons per km ²) |
|---------------|--|---------|---------|---------|---------|--------------|--|
| | 2016 WY | 2017 WY | 2018 WY | 2019 WY | 2020 WY | 2016–20 mean | 2016–20 mean |
| Vermilion | 455 | 300 | 390 | 344 | 317 | 361 | 108 |
| Embarras | 1,177 | 926 | 753 | 1,200 | 1,037 | 1,018 | 168 |
| Little Wabash | 643 | 454 | 492 | 858 | 743 | 638 | 80 |
| Rock | 791 | 1,101 | 678 | 891 | 661 | 824 | 33 |
| Green | 303 | 169 | 114 | 344 | 109 | 208 | 80 |
| Illinois | 4,771 | 3,754 | 3,315 | 4,629 | 4,092 | 4,112 | 59 |
| Kaskaskia | 955 | 556 | 745 | 978 | 1,228 | 892 | 66 |
| Big Muddy | 328 | 233 | 246 | 280 | 243 | 266 | 48 |
| Total | 9,421 | 7,494 | 6,733 | 9,523 | 8,430 | 8,320 | NA |

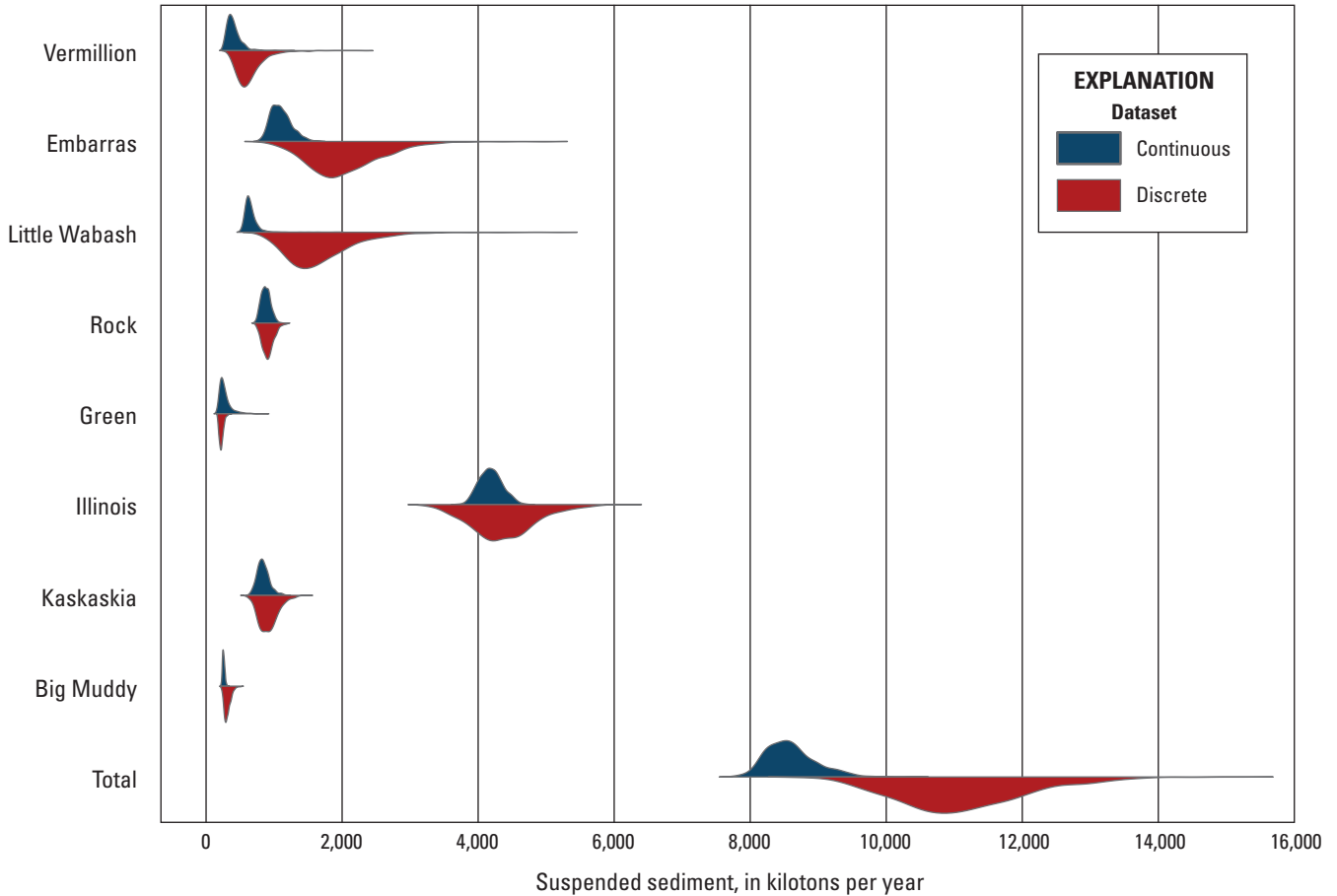


Figure 7. Graph showing mean annual suspended-sediment load by river based on continuous monitoring with Bayesian imputation (blue) and discrete sampling with C5-like model (red).

minimizes the importance and effects of regularization, which is another inherent benefit of continuous monitoring over discrete sampling.

There were several exceptions to the general trend of information gain scaling with streamflow, however. Continuously monitored surrogates that were strongly correlated with the target constituent added more information than weakly correlated surrogates. In general, optically measured NO₃ concentration was strongly correlated with true nitrate concentration, as was turbidity with suspended-sediment concentration, so these data typically added substantial information. Conversely, surrogates for TP were typically not as strongly correlated, so continuous monitoring tended to be less beneficial for estimating phosphorus loads. Furthermore, the strength of each surrogate varied among watersheds, because of differences in geology and land use.

These factors, as well as the extent of data gaps (table 3), all affect the information gain from continuous monitoring. For example, continuous monitoring of the Embarras River contributed a large amount of information, apparently because the discrete-data model performed unusually poorly at this station. In cases like this, preexisting sampling data could help inform the design of a new continuous monitoring network.

Where available, sampling data could be used to make preliminary assessments of load uncertainty to prioritize sites for continuous monitoring.

Using information gain to assess the value of continuous monitoring has an important shortcoming. These uncertainty estimates are biased because the models make unrealistic assumptions, mainly that model errors are uncorrelated in time. This bias tends to increase with weaker surrogates because weak surrogates typically have larger and more correlated errors. Because the discrete (C5-like) model uses only streamflow and time as surrogates, both of which are weak, uncertainty in the discrete model tends to be larger than that of the continuous models. Furthermore, violations of the model assumptions also bias other quantities, such as estimates of the mean. In other words, there is no guarantee inferences from a biased model will be correct, which can have ramifications for scientists and policymakers.

Network Improvements

Uncertainty information is also useful in prioritizing upgrades and maintenance in existing monitoring networks. Of the three constituents in this study—nitrate, phosphorus, and suspended sediment—continuous monitoring yielded the least uncertainty for nitrate (fig. 5), probably because the optical nitrate sensor was a strong surrogate—reasonably linear and low error with independent and identically distributed errors—for the true nitrate concentration. For most monitoring stations, further improvements, such as protection against debris or ice, might improve nitrate data coverage but are unlikely to have a substantive effect on the overall (network-wide) load estimate. Most of the uncertainty in overall nitrate load was from the Rock and Illinois Rivers (fig. 5). Data coverage for nitrate at the Illinois River station was 94 percent (table 3), so the potential for improvement is small. Data coverage was lower at the Rock River station (68 percent), so there is more potential for improvement if conditions at the site permit.

Though uncertainty in suspended-sediment loading was larger than for nitrate, continuous monitoring again offered clear benefit by reducing uncertainty relative to the load estimates on the basis of discrete sampling (fig. 7). Compared to nitrate, the larger uncertainty in suspended sediment is consistent with turbidity being a weaker surrogate for sediment in the sense that the correlation between turbidity and sediment concentration is both weaker and less linear than the correlation between optical nitrate and true nitrate. It is assumed the difference in uncertainty between these constituents primarily reflects this difference in surrogate strength and not the extent of data gaps because both sets of surrogates had similar data coverage. Therefore, advances in suspended-sediment sensors might lead to better sediment load estimates in the future, but better protection for the existing monitoring installations or making more frequent maintenance visits is unlikely to substantially reduce uncertainty.

Continuous monitoring arguably offered the least benefit for measuring phosphorus load but also the most potential for improvement once onsite TP analyzers become available. The quality of existing TP surrogates, such as turbidity and PO₄, varied among the monitoring stations. In some watersheds, turbidity was a good surrogate for TP, and further improvements to these stations may offer little benefit for load estimation. Just two stations—the Illinois River and the Kaskaskia River—were responsible for most of the uncertainty in the overall phosphorus load (fig. 6). In the future, upgrading these two stations with TP analyzers could substantially lower the overall uncertainty in phosphorus loading.

Summary

From water years 2016 through 2020, the U.S. Geological Survey operated continuous monitoring stations at eight major rivers in Illinois with a combined drainage area

covering approximately 74 percent of the land area of the State. Information from these sites could provide a benchmark from which to evaluate changes in nutrient and sediment loads from rivers throughout Illinois, which is critical for assessing the efficacy of nutrient and soil management practices that are being implemented to conserve topsoil and reduce the State's contribution of nutrients entering the Gulf of Mexico.

The use of continuous monitoring is growing as water-quality sensors become less expensive and more reliable, but this shift also brings new challenges. Continuous monitoring data contain gaps that must be filled before the data can be used for inference. Continuous monitoring is also more costly than periodic sampling, and that added expense needs to be justified by showing that continuous monitoring results in better estimates relative to what would be obtained through periodic sampling. This study presents a new method of filling gaps in continuous data that corrects for missing data bias that affected previous gap-filling methods. Results from this new method indicate that continuous monitoring yielded lower estimates of loads, with lower uncertainties, relative to those computed with periodic sampling over the same period.

An uncertainty assessment is critical for any measurement used in science or policymaking, but it can also be useful operationally. This study provides examples of how uncertainty information can be used to plan or upgrade existing monitoring networks in a cost-effective manner, thereby helping to reduce the cost of continuous monitoring while maximizing its benefit.

References Cited

- Alexander, R.B., Smith, R.A., Schwarz, G.E., Boyer, E.W., Nolan, J.V., and Brakebill, J.W., 2008, Difference in phosphorus and nitrogen delivery to the Gulf of Mexico from the Mississippi River Basin: *Environmental Science & Technology*, v. 42, no. 3, p. 822–830, accessed August 2020 at <https://pubs.acs.org/doi/10.1021/es0716103>.
- Berliner, L.M., 1996, Hierarchical Bayesian time series models, in Hanson K.M., and Silver R.N., eds., *Maximum entropy and Bayesian methods—Fundamental theories of physics*: Dordrecht, Netherlands, Springer, no. 1, v. 79, p. 15–22, accessed August 2020 at https://doi.org/10.1007/978-94-011-5430-7_3.
- Cohn, T.A., Caulder, D.L., Gilroy, E.J., Zynjuk, L.D., and Summers, R.M., 1992, The validity of a simple statistical model for estimating fluvial constituent loads—An empirical study involving nutrient loads entering Chesapeake Bay: *Water Resources Research*, v. 28, no. 9, p. 2353–2363, accessed August 2020 at <https://doi.org/10.1029/92WR01008>.

- David, M.B., Drinkwater, L.E., and McIsaac, G.F., 2010, Sources of nitrate yields in the Mississippi River Basin: *Journal of Environmental Quality*, v. 39, no. 5, p. 1657–1667, accessed August 2020 at <https://doi.org/10.2134/jeq2010.0115>.
- David, M.B., and Gentry, L.E., 2000, Anthropogenic inputs of nitrogen and phosphorus and riverine export for Illinois, USA: *Journal of Environmental Quality*, v. 29, no. 2, p. 494–508.
- David, M.B., McIsaac, G.F., Schnitkey, G.D., Czapar, G.F., and Mitchell, C.A., 2015, Science assessment, *in* Illinois Nutrient Loss Reduction Strategy: Illinois Environmental Protection Agency, p. 21–75, accessed August 2020 at <https://www2.illinois.gov/epa/Documents/iepa/water-quality/watershed-management/nlrs/nlrs-final-revised-083115.pdf>.
- Duda, A.M., 1985, Environmental and economic damage caused by sediment from agricultural nonpoint sources: *Journal of the American Water Resources Association*, v. 21, no. 2, p. 225–234, accessed August 2021 at <https://doi.org/10.1111/j.1752-1688.1985.tb00132.x>.
- Fishman, M.J., 1993, Methods of analysis by the U.S. Geological Survey National Water Quality Laboratory—Determination of inorganic and organic constituents in water and fluvial sediments: U.S. Geological Survey Open-File Report 93-125, 217 p., accessed August 2021 at <https://doi.org/10.3133/ofr93125>.
- Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., and Rubin, D.B., 2013, Bayesian data analysis: Boca Raton, Fla., Chapman and Hall/CRC, 639 p.
- Gelman, A., Jakulin, A., Pittau, M.G., and Su, Y.S., 2008, A weakly informative default prior distribution for logistic and other regression models: *The Annals of Applied Statistics*, v. 2, no. 4, p. 1360–1383, accessed August 2020 at <https://doi.org/10.1214/08-AOAS191>.
- Heimann, D.C., Sprague, L.A., and Blevins, D.W., 2011, Trends in suspended-sediment loads and concentrations in the Mississippi River Basin: Scientific Investigation Report 2011-5200, 33 p., accessed August 2020 at <https://doi.org/10.3133/sir20115200>.
- Hodson, T.O., Terrio, P.J., Peake, C.S., and Fazio, D.J., 2021, Modeled nutrient and sediment concentrations from major rivers in Illinois based on continuous monitoring from October 1, 2015, through September 30, 2020: U.S. Geological Survey data release, <https://doi.org/10.5066/P9MS80IH>.
- Illinois Environmental Protection Agency, Illinois Department of Agriculture, and University of Illinois Extension Illinois, 2015, Illinois Nutrient Loss Reduction Strategy, accessed August 2020 at <https://www2.illinois.gov/epa/Documents/iepa/water-quality/watershed-management/nlrs/nlrs-final-revised-083115.pdf>.
- Jacobson, L.M., David, M.B., and Drinkwater, L.E., 2011, A spatial analysis of phosphorus in the Mississippi River Basin: *Journal of Environmental Quality*, v. 40, no. 3, p. 931–941, accessed August 2020 <https://doi.org/10.2134/jeq2010.0386>.
- Kennedy, E.J., 1983, Computation of continuous records of streamflow: Techniques of Water-Resources Investigations of the United States Geological Survey, 53 p, accessed August 2020 at <https://doi.org/10.3133/twri03A13>.
- Lathrop, T.R., Bunch, A.R., and Downhour, M.S., 2019, Regression models for estimating sediment and nutrient concentrations and loads at the Kankakee River, Shelby, Indiana, December 2015 through May 2018: U.S. Geological Survey Scientific Investigation Report 2019–5005, 13 p, accessed August 2020 at <https://doi.org/10.3133/sir20195005>.
- Lewandowski, D., Kurowicka, D., and Joe, H., 2009, Generating random correlation matrices based on vines and extended onion method: *Journal of Multivariate Analysis*, v. 100, no. 9, p. 1989–2001, accessed August 2020 <https://doi.org/10.1016/j.jmva.2009.04.008>.
- Little, R.J.A., and Rubin, D.B., 2019, Statistical analysis with missing data 3rd ed.: New York, John Wiley & Sons, 464 p.
- Mueller, D.S., Wagner, C.R., Rehm, M.S., Oberg, K.A., and Rainville, F., 2013, Measuring discharge with acoustic Doppler current profilers from a moving boat: U.S. Geological Survey Techniques and Methods, book 3, chap. A22, 95 p.
- Patton, C.J., and Kryskalla, J.R., 2003, Methods of analysis by the U.S. Geological Survey National Water Quality Laboratory—Evaluation of alkaline persulfate digestion as an alternative to Kjeldahl digestion for determination of total and dissolved nitrogen and phosphorus in water: U.S. Geological Survey Water-Resources Investigations Report 2003-4174, accessed August 2021 at <https://doi.org/10.3133/wri034174>.
- Patton, C.J., and Kryskalla, J.R., 2011, Colorimetric determination of nitrate plus nitrite in water by enzymatic reduction, automated discrete analyzer methods: U.S. Geological Survey Techniques and Methods, book 5, chap. B8, accessed August 2021 at <https://doi.org/10.3133/tm5B8>.

- Robertson, D.M., Hubbard, L.E., Lorenz, D.L., and Sullivan, D.J., 2018, A surrogate regression approach for computing continuous loads for the tributary nutrient and sediment monitoring program on the Great Lakes: *Journal of Great Lakes Research*, v. 44, no. 1, p. 26–42, accessed August 2020 <https://doi.org/10.1016/j.jglr.2017.10.003>.
- Robertson, D.M., Saad, D.A., and Schwarz, G.E., 2014, Spatial variability in nutrient transport by HUC8, State, and subbasin based on Mississippi/Atchafalaya River Basin SPARROW models: *Journal of the American Water Resources Association*, v. 50, no. 4, p. 988–1009, accessed August 2020 at <https://doi.org/10.3133/tm3A22>.
- Robertson, D.M., and Saad, D.A., 2019, Spatially referenced models of streamflow and nitrogen, phosphorus, and suspended-sediment loads in streams of the Midwestern United States: U.S. Geological Survey Scientific Investigations Report 2019–5114, 74 p. including 5 appendices, accessed August 2020 at <https://doi.org/10.3133/sir20195114>.
- Salvatier, J., Wiecki, T. V., and Fonnesbeck, C., 2016, Probabilistic programming in Python using PyMC3: *PeerJ Computer Science*, v. 2016, no. 4, p. 1–24, accessed August 2020 at <https://doi.org/10.7717/peerj-cs.55>.
- Sprague, L.A., Hirsch, R.M., and Aulenbach, B.T., 2011, Nitrate in the Mississippi river and its tributaries, 1980 to 2008—Are we making progress?: *Environmental Science & Technology*, v. 45, no. 17, p. 7209–7216, accessed August 2020 <https://doi.org/10.1021/es201221s>.
- Terrio, P.J., Straub, T.D., Domanski, M.M., and Siudyla, N.A., 2015, Continuous monitoring of sediment and nutrients in the Illinois River at Florence, Illinois, 2012–13: U.S. Geological Survey Scientific Investigations Report 2015–5040, 61 p., accessed August 2020 at <http://dx.doi.org/10.3133/sir20155040>.
- Turnipseed, D.P., and Sauer, V.B., 2010, Discharge measurements at gaging stations: U.S. Geological Survey Techniques and Methods, book 3, chap. A8, 87 p, accessed August 2020 at <https://doi.org/10.3133/tm3A8>.
- U.S. Geological Survey, 2021, USGS water data for the Nation: U.S. Geological Survey National Water Information System database, accessed January 2021 at <https://doi.org/10.5066/F7P55KJN>.
- Wagner, R.J., Boulger, R.W., Jr., Oblinger, C.J., and Smith, B.A., 2006, Guidelines and standard procedures for continuous water-quality monitors—Station operation, record computation, and data reporting: U.S. Geological Survey Techniques and Methods 1–D3, 51 p., 8 attachments, accessed August 2020 at <https://doi.org/10.3133/tm1D3>.

Appendix 1. Station Descriptions

This appendix describes summary statistics, seasonality, diurnal cycling, and storm-event patterns for each parameter during the study period, as well as descriptions of setting, instrumentation, and streamflow characteristics for each station. Summary statistics for each parameter are shown by site in [table 1.1](#). These statistics were based on 15-minute measurements, and not daily means. Furthermore, the statistics are based only on the observed data record that contains substantial gaps ([table 3](#)) that may bias these estimates compared to the true range of conditions.

Seasonality

Water temperature fluctuated seasonally with air temperature at all sites. Seasonal patterns in specific conductance were site dependent. In the Vermilion, Rock, and Illinois Rivers, specific conductance peaked during the winter, whereas in the Big Muddy River, specific conductance peaked in the fall. Specific conductance did not have strong seasonality in the Embarras, Little Wabash, Green, or Kaskaskia Rivers. Dissolved oxygen varied seasonally at all sites with concentrations peaking during winter. Seasonality in dissolved oxygen concentrations was typically larger in smaller watersheds. pH did not vary seasonally in its mean, but it did in its variance with typically less variability in winter. Seasonally, turbidity did not have many patterns in magnitude or frequency except during summer and early fall. During these times, streamflows were typically lower resulting in less turbid conditions at all sites. All rivers, except the Big Muddy River, had seasonality in nitrate concentrations with higher concentrations in late spring or early summer and lower concentrations in late summer or early fall.

Diurnal Cycling

Diurnal cycling was observed in most watersheds. Water temperature cycled diurnally at moderate to low streamflows, with more pronounced diurnal cycling during the summer. Specific conductance generally did not exhibit diurnal cycling

except at the Rock River where it occurred during periods of steady streamflow. Dissolved oxygen cycled diurnally at all sites, with the largest fluctuations occurring at low streamflow and on smaller rivers. Diurnal cycling of pH was more pronounced during the summer and in smaller watersheds. Diurnal cycling of turbidity occurred at the Embarras, Little Wabash, Green, Illinois, Kaskaskia and Big Muddy Rivers and was most pronounced during summer low streamflows. Diurnal cycling was not observed in nitrate; if it occurred, it was below the resolution of the sensors used in this study.

Storm Events

For most constituents, storm events disrupted the general seasonal and diurnal patterns. During rain events, water temperatures typically rose during winter and fell during summer at all sites. Specific conductance would decrease rapidly during storm events, then gradually return to pre-event levels as streamflow decreased. Dissolved oxygen typically decreased during storm events regardless of the season, although sometimes it increased. pH typically decreased during storm events, and the magnitude of decrease correlated with the magnitude of the storm event. pH gradually returned to pre-event conditions as streamflow decreased. Turbidity and streamflow were correlated at all sites; however, turbidity is influenced by other factors such that the largest streamflows were not always associated with the highest turbidities. The highest turbidities occurred as streamflow increased during storm events. The response of nitrate concentrations during storm events varied among sites. The general pattern is that the initial streamflow increase from a rain event will reduce nitrate concentrations, but as streamflows decrease, nitrate will peak. This general pattern is applicable outside of the spring and early summer window when concentrations are highest. During the spring to early summer period, nitrate concentrations' response to rain events is highly erratic. In this period, rain events can cause increases or decreases in nitrate concentrations, as well as differences in lag times between peak streamflows and peak nitrate concentrations.

Table 1.1. Continuous data summary by constituent and river for the 2016–20 water years. Data available from the National Water Information System (U.S. Geological Survey, 2021).[°C, degrees Celsius; $\mu\text{S}/\text{cm}$, microsiemens per centimeter; mg/L , milligrams per liter; FNU, Formazin Nephelometric Units; $\text{mg P}/\text{L}$, milligrams of phosphorus per liter; m^3/s , cubic meters per second]

| Constituent | River | Minimum | Maximum | Median | 25th percentile | 75th percentile |
|--|---------------|---------|---------|--------|-----------------|-----------------|
| Temperature (°C) | Vermilion | 0.8 | 32.2 | 17.1 | 9.1 | 24.5 |
| | Embarras | -0.1 | 32.7 | 15 | 6.9 | 24.5 |
| | Little Wabash | 0.1 | 31.8 | 17.4 | 7.1 | 24.8 |
| | Rock | -0.1 | 30 | 15.5 | 6.5 | 23.4 |
| | Green | -0.1 | 30.4 | 15.1 | 7.4 | 21.8 |
| | Illinois | 0 | 31.8 | 14 | 5.5 | 24.7 |
| | Kaskaskia | 0 | 32.6 | 16 | 7.1 | 24.8 |
| | Big Muddy | -0.1 | 31.3 | 14.9 | 7.3 | 24.7 |
| Specific conductance ($\mu\text{S}/\text{cm}$) | Vermilion | 237 | 918 | 598 | 546 | 638 |
| | Embarras | 104 | 732 | 449 | 344 | 534 |
| | Little Wabash | 100 | 825 | 292 | 208 | 436 |
| | Rock | 349 | 883 | 645 | 599 | 683 |
| | Green | 231 | 901 | 690 | 658 | 714 |
| | Illinois | 389 | 1040 | 698 | 628 | 771 |
| | Kaskaskia | 112 | 682 | 357 | 304 | 419 |
| | Big Muddy | 118 | 2020 | 516 | 410 | 675 |
| Dissolved oxygen (mg/L) | Vermilion | 4.3 | 20.9 | 9.6 | 8 | 11.6 |
| | Embarras | 0.2 | 20.4 | 9.2 | 7 | 11.3 |
| | Little Wabash | -0.1 | 13.4 | 6.25 | 4.8 | 8.9 |
| | Rock | 4.8 | 18.7 | 10.4 | 8.4 | 13 |
| | Green | 2.6 | 16.2 | 9.6 | 8 | 11.4 |
| | Illinois | 2.7 | 16 | 9.6 | 6.4 | 12.1 |
| | Kaskaskia | -0.1 | 21.4 | 8.5 | 6 | 11.4 |
| | Big Muddy | -0.1 | 14.1 | 7.4 | 4.7 | 10 |
| pH | Vermilion | 7.4 | 9.1 | 8.2 | 8.1 | 8.4 |
| | Embarras | 6.7 | 9 | 7.9 | 7.7 | 8.2 |
| | Little Wabash | 6.4 | 8.4 | 7.3 | 7.1 | 7.6 |
| | Rock | 7.5 | 9 | 8.4 | 8.2 | 8.5 |
| | Green | 7 | 8.7 | 8 | 7.8 | 8.1 |
| | Illinois | 7.3 | 9.3 | 8.1 | 7.9 | 8.3 |
| | Kaskaskia | 6.6 | 9 | 7.7 | 7.5 | 7.9 |
| | Big Muddy | 6.7 | 8.5 | 7.2 | 7.1 | 7.4 |
| Turbidity hach (FNU) | Vermilion | 2.9 | 1820 | 18.1 | 10 | 43.8 |
| | Embarras | 2.6 | 1780 | 46 | 21 | 128 |
| | Little Wabash | 3.9 | 3940 | 80 | 43.1 | 140 |
| | Rock | 0.0 | 861 | 28.4 | 18 | 43.2 |
| | Green | 0.5 | 1990 | 12.9 | 7.5 | 26 |
| | Illinois | 5.9 | 1400 | 61 | 39 | 103 |
| | Kaskaskia | 2.7 | 759 | 40 | 24 | 78.6 |
| | Big Muddy | 3.6 | 1140 | 45.5 | 28.8 | 72 |

Table 1.1. Continuous data summary by constituent and river for the 2016–20 water years. Data available from the National Water Information System (U.S. Geological Survey, 2021).—Continued

[°C, degrees Celsius; $\mu\text{S}/\text{cm}$, microsiemens per centimeter; mg/L , milligrams per liter; FNU, Formazin Nephelometric Units; $\text{mg P}/\text{L}$, milligrams of phosphorus per liter; m^3/s , cubic meters per second]

| Constituent | River | Minimum | Maximum | Median | 25th percentile | 75th percentile |
|---|---------------|---------|---------|--------|-----------------|-----------------|
| Turbidity YSI (FNU) | Vermilion | 0 | 1300 | 16.3 | 10 | 35.3 |
| | Embarras | 0.3 | 991 | 33.7 | 17.8 | 94.2 |
| | Little Wabash | 0.4 | 1580 | 58.1 | 30.6 | 104 |
| | Rock | 1.4 | 561 | 25.2 | 15 | 37.4 |
| | Green | 0 | 1290 | 12.4 | 8.2 | 27.7 |
| | Illinois | 0.1 | 1630 | 45.9 | 30.1 | 77.5 |
| | Kaskaskia | 1.5 | 540 | 34 | 21 | 64.9 |
| | Big Muddy | 2 | 758 | 35.8 | 22.4 | 59.2 |
| Nitrate plus nitrite (mg/L) | Vermilion | 0.3 | 25.1 | 5.7 | 3.75 | 7.31 |
| | Embarras | 0 | 14.1 | 2.32 | 0.8 | 4 |
| | Little Wabash | 0.1 | 7 | 1.2 | 0.8 | 1.7 |
| | Rock | 1.4 | 13.5 | 4.1 | 3.57 | 5.11 |
| | Green | 0.5 | 13.7 | 4.1 | 3.1 | 5.58 |
| | Illinois | 0.8 | 7.51 | 3.74 | 2.79 | 4.6 |
| | Kaskaskia | 0 | 9.43 | 0.9 | 0.5 | 1.5 |
| | Big Muddy | 0 | 4.79 | 0.5 | 0.39 | 0.7 |
| Orthophosphate ($\text{mg P}/\text{L}$) | Vermilion | 0.05 | 1.37 | 0.19 | 0.1 | 0.23 |
| | Embarras | 0.001 | 0.777 | 0.09 | 0.05 | 0.21 |
| | Little Wabash | 0.01 | 0.287 | 0.1 | 0.06 | 0.14 |
| | Rock | 0.003 | 0.163 | 0.07 | 0.05 | 0.09 |
| | Green | 0.004 | 0.356 | 0.03 | 0.02 | 0.04 |
| | Illinois | 0.075 | 0.512 | 0.22 | 0.18 | 0.27 |
| | Kaskaskia | 0.002 | 0.621 | 0.17 | 0.11 | 0.23 |
| | Big Muddy | 0.005 | 0.458 | 0.12 | 0.08 | 0.16 |
| Streamflow (m^3/s) | Vermilion | 1.1 | 1124.2 | 22.9 | 9.3 | 43.9 |
| | Embarras | 0.8 | 1067.5 | 33.7 | 12.1 | 115.0 |
| | Little Wabash | 0.4 | 900.5 | 62.3 | 10.3 | 211.0 |
| | Rock | 90.6 | 1257.3 | 328.5 | 223.1 | 470.1 |
| | Green | 3.1 | 322.8 | 24.7 | 15.2 | 38.8 |
| | Illinois | 90.6 | 3284.7 | 821.2 | 475.7 | 1214.8 |
| | Kaskaskia | 1.4 | 1565.9 | 122.9 | 46.2 | 286.0 |
| | Big Muddy | -79.3 | 903.3 | 40.2 | 16.2 | 100.0 |

Vermilion River near Danville, Illinois

U.S. Geological Survey (USGS) station number 03339000 (Illinois Environmental Protection Agency [IEPA] station number BP-08)

Location—Latitude 40°06'03", longitude 87°35'50", referenced to North American Datum of 1983, in NW¼NW¼ sec. 22, T. 19 N., R. 11 W., Vermilion County, Ill., hydrologic unit 05120109, on right bank at Danville sewage-treatment

plant, 1.7 miles upstream from Stony Creek, 2.2 miles south-east of Danville, and at mile 19.5 (U.S. Geological Survey, 2021).

Station description—The river is entrenched at this site, promoting high velocities and streamflows that transport abundant debris and sediment. The bridge and channel configuration prevents a bridge-pier installation, so the sensors were initially on the streambank; however, the bank experienced rapid sediment deposition during storm events. Consequently, the sensors were relocated to a gage house with a pumping

system atop the streambank in the summer of 2017. Since that time, the longest data gaps are related to low water levels and damage sustained during the February 2018 flood event. Shortly after the pumping system was installed, the WET Labs Cycle orthophosphate (Cycle-PO₄) analyzer was replaced with a YSI P700 orthophosphate (PO₄) analyzer, which was better suited to the high PO₄ and sediment concentrations at the site. Before the installation of the gage house, the sensors were removed during the winter because of ice accumulation. At very low streamflows, the pump intake can come out of the water. During periods of extended drought (2019), this can result in months of missing data. Although drought periods do not contribute substantial nutrient loads, they can impact other designated uses, such as supporting aquatic life and recreation, so these gaps may negatively impact some secondary monitoring objectives. During the spring of 2018, an autosampler was installed to sample high-turbidity events.

Instrumentation—The station includes the following water-quality monitoring instrumentation:

- YSI EXO 2 with water temperature, specific conductance, dissolved oxygen, pH, and turbidity probes measuring at a 15-minute interval;
- Hach Nitratex nitrate sensor measuring at a 15-minute interval;
- Hach Solitax turbidity sensor measuring at a 15-minute interval;
- Sea-Bird Cycle-PO₄ PO₄ analyzer measuring at a 2-hour interval (discontinued in 2017);
- YSI P700 PO₄ analyzer measuring at a 15-minute interval;
- ISCO refrigerated autosampler triggered by turbidity; and a
- Campbell Scientific CR6 datalogger.

Streamflow characteristics—During the study, this site experienced two of the largest streamflow events on record (fig. 1.1): the first, in December 2015, was the 6th largest recorded streamflow event with daily mean streamflow at slightly under 991 cubic meters per second (m³/s), and the second, in February 2018, was the third largest recorded streamflow event with a mean streamflow of about 1,050 m³/s. The largest recorded streamflow was 1,380 m³/s in March 1939. Extended periods of low streamflows are common during late summer and early fall. The lowest daily mean streamflow during the study was 1.3 m³/s in September 2019, with the lowest on record being 0.05 m³/s in 1920 and 1930.

Embarras River at Lawrenceville, Ill

USGS station number 03346500 (IEPA site number BE-01)

Location—Latitude 38°43'25", longitude 87°39'52", referenced to North American Datum of 1983, in NE¼SW¼ sec. 5, T. 3 N., R. 11 W., Lawrence County, Ill., hydrologic unit 05120112, on left bank at downstream side of U.S. Business Route 50 bridge in Lawrenceville, and at mile 6.7 (U.S. Geological Survey, 2021).

Station description—Because of the configuration of the stream channel and bridge piers, the sensors at this station are positioned in perforated pipes mounted to the stream-channel side of a bridge pier. The sensors are serviced from a platform mounted to the pier and are inaccessible above moderate stages. PO₄ monitoring was discontinued in 2018 because of the excessive fouling caused by high turbidity levels at this site.

Instrumentation—The station includes the following water-quality monitoring instrumentation:

- YSI EXO 2 with water temperature, specific conductance, dissolved oxygen, pH, and turbidity probes measuring at a 15-minute interval;
- Hach Nitratex nitrate sensor measuring at a 15-minute interval;
- Hach Solitax turbidity sensor measuring at a 15-minute interval;
- Sea-Bird Cycle-PO₄ PO₄ analyzer measuring at a 2-hour interval (discontinued in 2018); and a
- Campbell Scientific CR6 datalogger.

Streamflow characteristics—The highest streamflow during the study period occurred in early May 2017, with a mean daily streamflow of 1,056 m³/s (fig. 1.2). This was the 5th largest mean daily streamflow on record, with the highest recorded streamflow of 1,345 m³/s recorded in June 2008. Later, in October 2017, the lowest mean daily streamflow was observed at only 0.97 m³/s, with the lowest recorded mean daily streamflow of 0.71 m³/s recorded in August 2012. The typical pattern of higher streamflows after spring rain events and low streamflows in the late summer/early fall predominates during the study period and historical record.

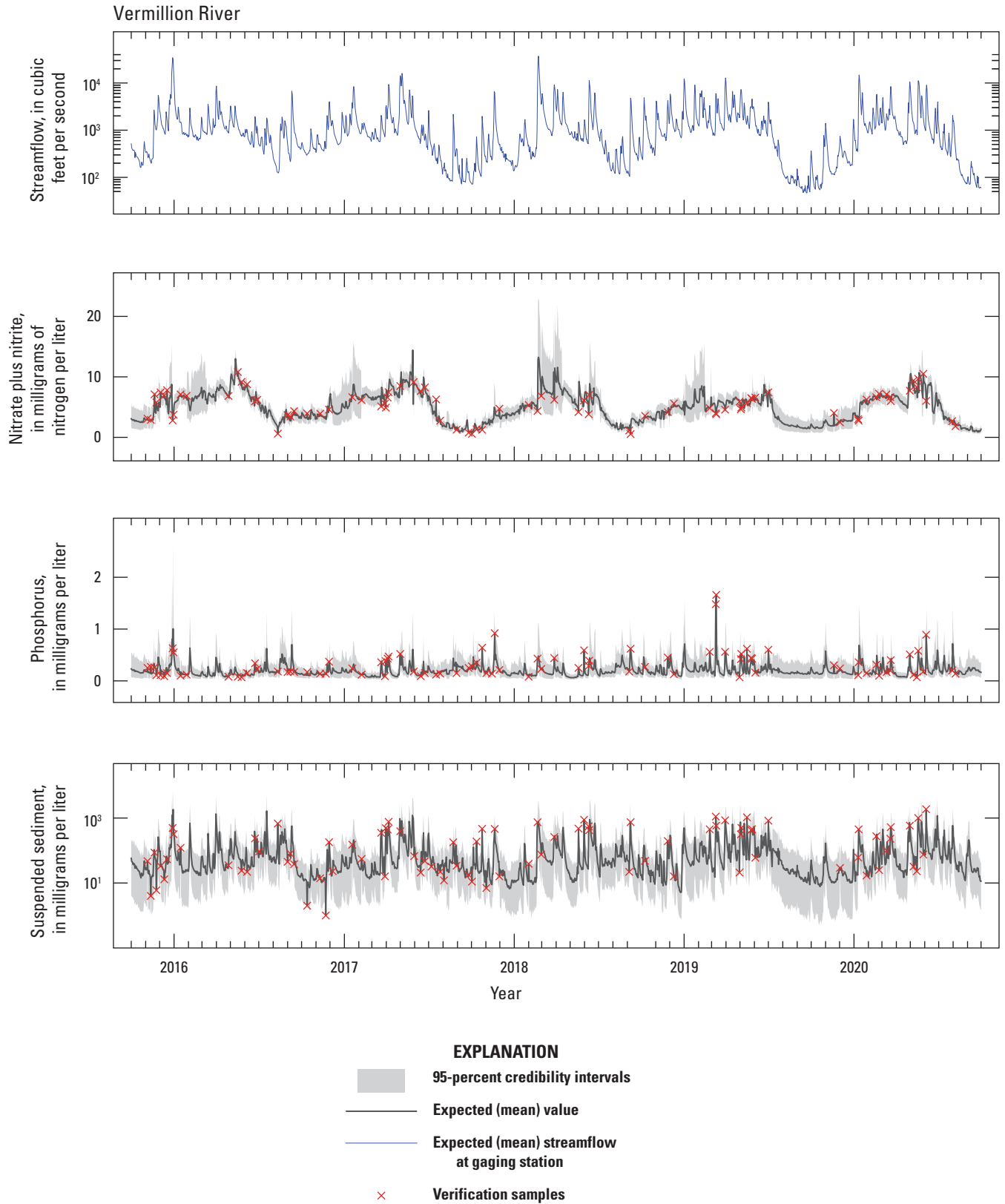


Figure 1.1. Timeseries of streamflow, nitrate plus nitrite, phosphorus, and suspended sediment for the Vermilion River at Danville.

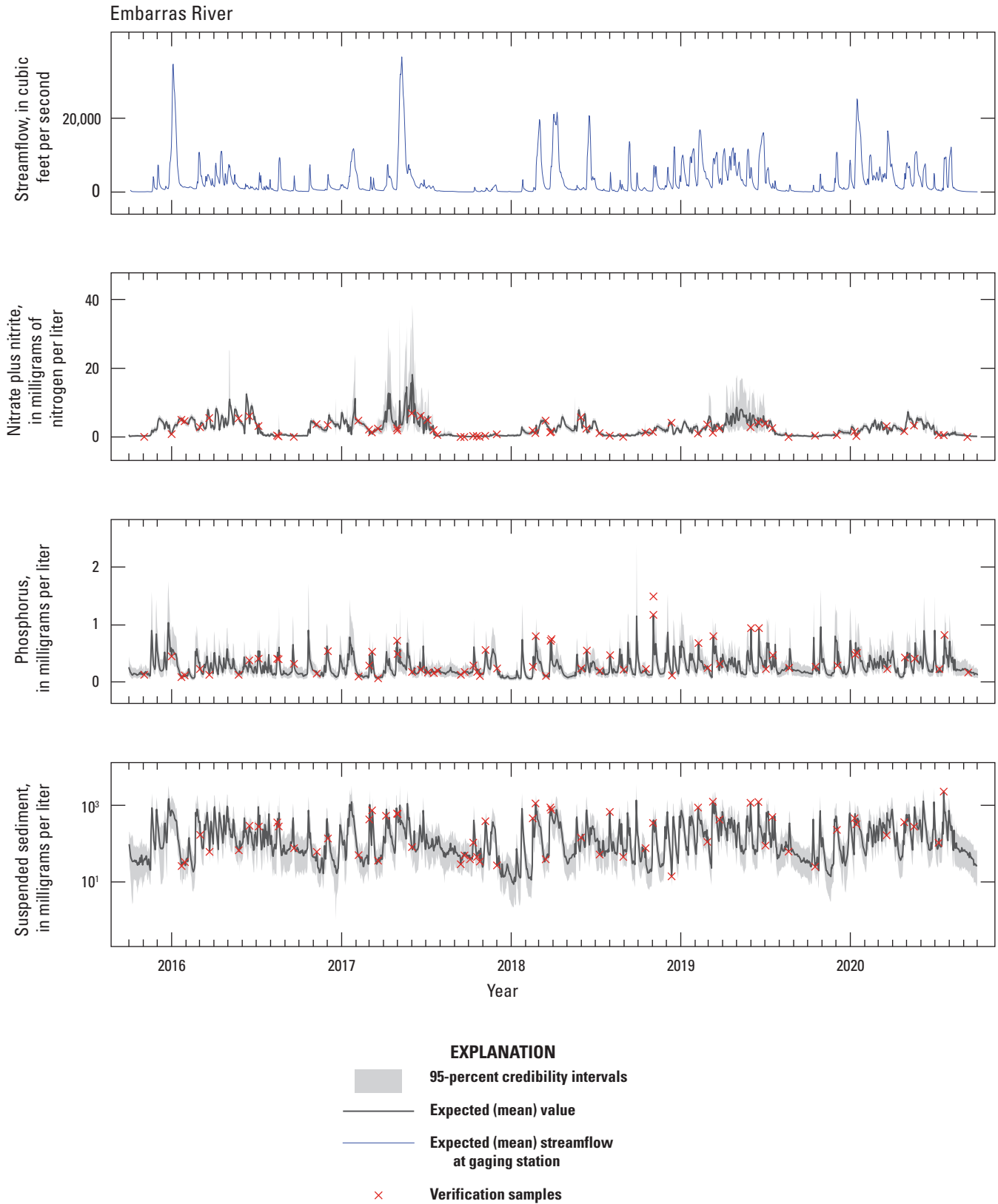


Figure 12. Timeseries of streamflow, nitrate plus nitrite, phosphorus, and suspended sediment for the Embarras River at Lawrenceville.

Little Wabash River at Main Street at Carmi, Ill

USGS station number 03381495 ([IEPA] site number C-23)

Location—Latitude 38°05'32", longitude 88°09'22" referenced to North American Datum of 1983, in NW¼SE¼NE¼ sec. 13, T. 5 S., R. 9 E., White County, Ill., hydrologic unit 05120114, at auxiliary gaging station at Main Street Bridge in Carmi, 5.0 miles downstream from Skillet Fork and at mile 33.8 (U.S. Geological Survey, 2021).

Station description—This station presents some challenges because of the large range of stream stages throughout the year. Additionally, a massive debris jam on the upstream side of the bridge caused streambed conditions to change rapidly. The pipes that house the sensors have been modified several times to accommodate variable sedimentation rates. The sensors are mounted as near to the streambed as practical, but they come out of water during low streamflow, which is detrimental to some instruments. As a result, the YSI EXO 2 is removed during low streamflows, but the Hach Nitratax and Solitax are left in place. The Hach sensors are less likely to be damaged when running dry, and this allows nitrate and turbidity data to be recorded above a certain stage when field servicing may not be practical. PO₄ monitoring was discontinued here because of the high turbidity levels and the sensors coming out of the water. Also, fish kills occurred near the gage in at least 3 of the 5 years during the study period.

Instrumentation—The station includes the following water-quality monitoring instrumentation:

- YSI EXO 2 with water temperature, specific conductance, dissolved oxygen, pH, and turbidity probes measuring at a 15-minute interval;
- Hach Nitratax nitrate sensor measuring at a 15-minute interval;
- Hach Solitax turbidity sensor measuring at a 15-minute interval;
- Sea-Bird Cycle-PO₄ PO₄ analyzer measuring at a 2-hour interval (discontinued in 2017); and a
- Campbell Scientific CR6 datalogger.

Streamflow characteristics—Streamflow during the study period was largely normal with a range of streamflows consistent with historical events since the site's inception in 1992. The Little Wabash River at Main Street at Carmi is the auxiliary gage to the primary gage, Little Wabash River at Carmi (USGS station number 03381500), where streamflow is calculated using the water surface slope between the two gages. Because no major inflows between the gages are present, streamflow from the primary gage can be used. During the period of record, the highest mean daily streamflow was approximately 900 m³/s on May 8, 2017 (fig. 1.3), with the greatest recorded mean daily streamflow of 1,565 m³/s in May 2011. The lowest streamflow occurred in October 2017 with a mean daily streamflow of 0.4 m³/s, with the lowest recorded being 0.017 m³/s in July 1954. The typical pattern of higher streamflows after spring rain events and low streamflows in the late summer/early fall predominates during the study period and historical record.

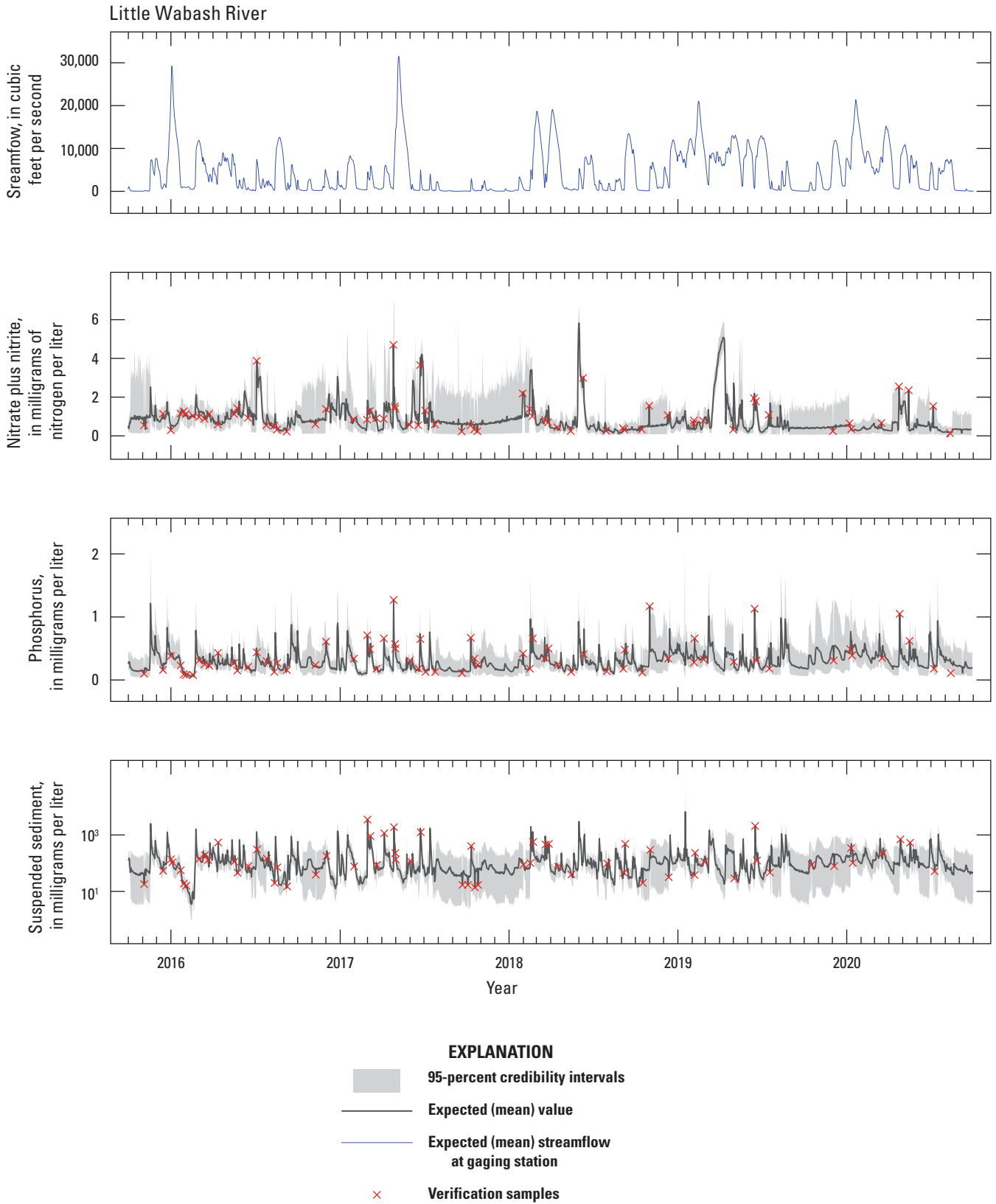


Figure 1.3. Timeseries of streamflow, nitrate plus nitrite, phosphorus, and suspended sediment for the Little Wabash River at Carmi.

Rock River near Joslin, Ill

USGS station number 05446500 (IEPA site number P-04)

Location—Latitude 41°33'22", longitude 90°11'07" referenced to North American Datum of 1983, in NE¼NE¼ sec. 18, T. 18 N., R. 3 E., Rock Island County, Ill., hydrologic unit 07090005, near center of span on downstream side of bridge on State Highway 92, 1.8 miles east of Joslin, 14.5 miles downstream from Rock Creek, and at mile 26.9 (U.S. Geological Survey, 2021).

Station description—Maintaining infrastructure and data record at this station was problematic. The instrumentation was initially installed in heavy-gauge polyvinyl chloride conduits mounted to a large metal platform installed to girders on the downstream side of a bridge pier near the center of the channel. However, seasonal ice formation and ice flows damaged the conduits and mounts every winter, so it was not practical to keep the instrumentation installed during the winter. Sometimes there were substantial periods between when damage occurred and when repairs could be made, because of safety concerns and streamflow conditions limiting access to the damaged areas. In 2019, the fixed conduits were replaced with swinging, chain-suspended conduits allowing the conduit positions to compensate for changes in streamflow and to ride up on top of ice flows in the hope that this might help mitigate some of the damage to the conduits and instrumentation. PO₄ monitoring was discontinued at this site because calcium carbonate accumulation damaged the analyzers. In 2018, an ISCO automated sampler was installed.

Instrumentation—The station includes the following water-quality monitoring instrumentation:

- YSI EXO 2 with water temperature, specific conductance, dissolved oxygen, pH, and turbidity probes measuring at a 15-minute interval;
- Hach Nitratax nitrate sensor measuring at a 15-minute interval;
- Hach Solitax turbidity sensor measuring at a 15-minute interval;
- Sea-Bird Cycle-PO₄ PO₄ analyzer measuring at a 2-hour interval (discontinued in 2017);
- ISCO autosampler triggered by turbidity; and a
- Campbell Scientific CR6 datalogger.

Streamflow characteristics—Streamflow during the study period was representative of the historical record. During the study period, the highest daily mean streamflow during the study was about 1,240 m³/s in February 2019 (fig. 1.4), with the highest daily mean streamflow recorded being 1,265 m³/s in June 2002. The lowest daily mean streamflow during the period was approximately 93 m³/s in October 2015, where the lowest daily mean streamflow recorded was 23.6 m³/s in January 1940 because of backwater effects from ice. The river is generally homogeneous at this location, and the monitoring station is well located in the middle of a straight section of the channel. Typically, there is substantial streamflow in the river throughout the year, with the highest streamflows during spring rain events and the lowest streamflows during the late summer and early fall months.

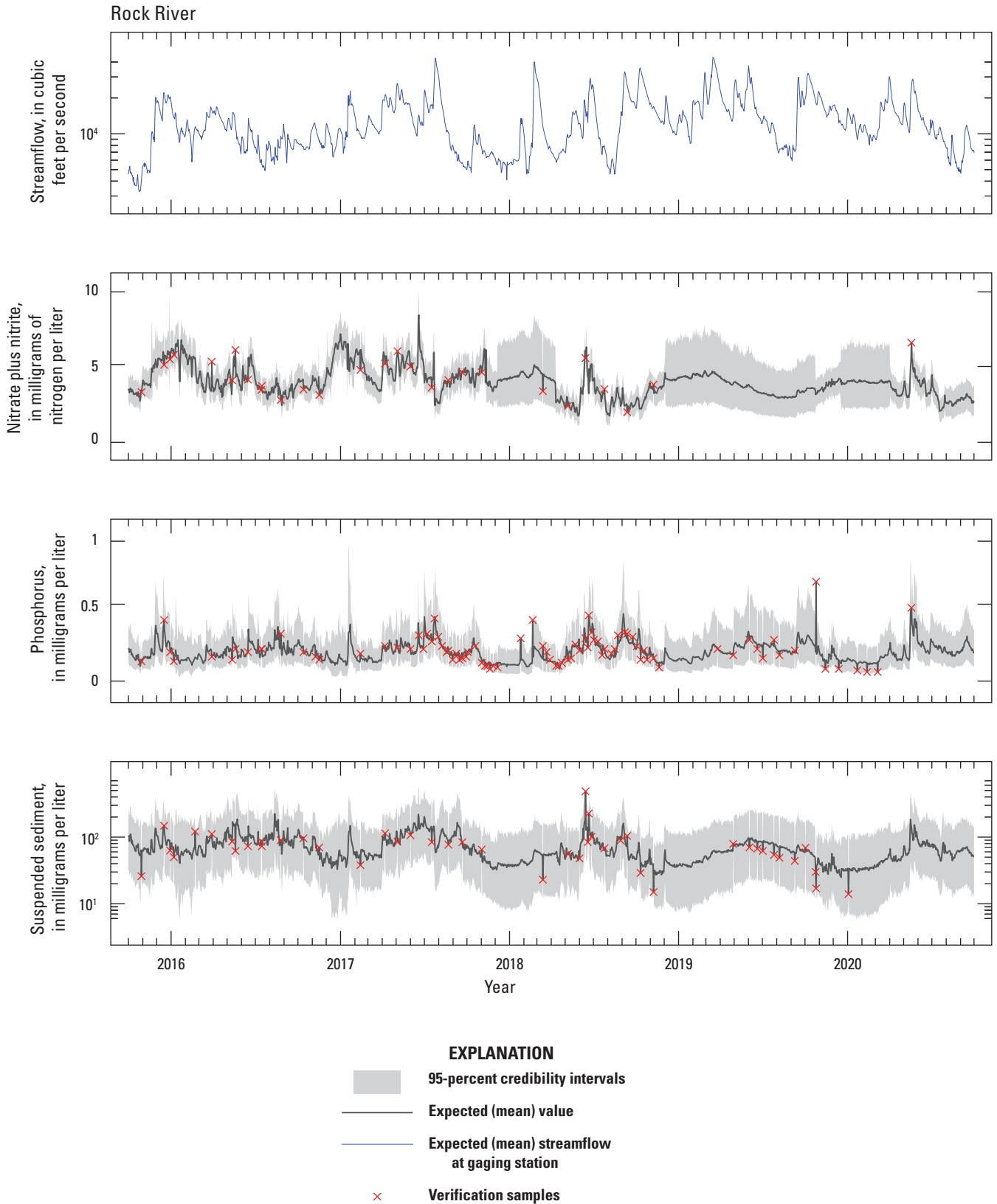


Figure 1.4. Timeseries of streamflow, nitrate plus nitrite, phosphorus, and suspended sediment for the Rock River at Joslin.

Green River near Geneseo, Ill

USGS station number 05447500 (IEPA site number PB-04)

Location—Latitude 41°29'20", longitude 90°09'27" referenced to North American Datum of 1983, in NE¼SW¼ sec. 4, T. 17 N., R. 3 E., Henry County, Ill., hydrologic unit 07090007, on upstream side of bridge on State Highway 82, 1.4 mi upstream from Geneseo Creek, 2.4 mi north of Geneseo, and at mile 14.9 (U.S. Geological Survey, 2021).

Station description—Because of the configuration of the stream channel and bridge piers, the sensors at this station are positioned in perforated conduits mounted to the stream-channel side of an abandoned bridge abutment. Because of ice flows, the sensors are routinely removed in the winter. Additionally, ice flows damaged the conduit resulting in extended intervals of missing data while repairs were made. PO4 monitoring was discontinued in 2017 because of high concentrations at or near the top of the analyzers range and the unreliability of the analyzer during high turbidity events. In 2018, an ISCO automated sampler was added.

Instrumentation—The station includes the following water-quality monitoring instrumentation:

- YSI EXO 2 with water temperature, specific conductance, dissolved oxygen, pH, and turbidity probes measuring at a 15-minute interval;
- Hach Nitratax nitrate sensor measuring at a 15-minute interval;
- Hach Solitax turbidity sensor measuring at a 15-minute interval;
- Sea-Bird Cycle-PO4 PO4 analyzer measuring at a 2-hour interval (discontinued in 2017);
- ISCO autosampler triggered by turbidity; and a
- Campbell Scientific CR6 datalogger.

Streamflow characteristics—Streamflow over the described study period was generally representative of the site since recording began in 1940. The highest streamflow during the study period occurred in early May 2019, with a mean daily streamflow of 300 m³/s (fig. 1.5). The highest mean daily streamflow recorded is 342 m³/s in June 1974. The lowest mean daily streamflow was observed at 4.4 m³/s in September 2017, with the lowest recorded mean daily streamflow of 0.6 m³/s recorded in January 1977. The typical pattern of higher streamflows after spring rain events and low streamflows in the late summer/early fall predominates during the study period and historical record.

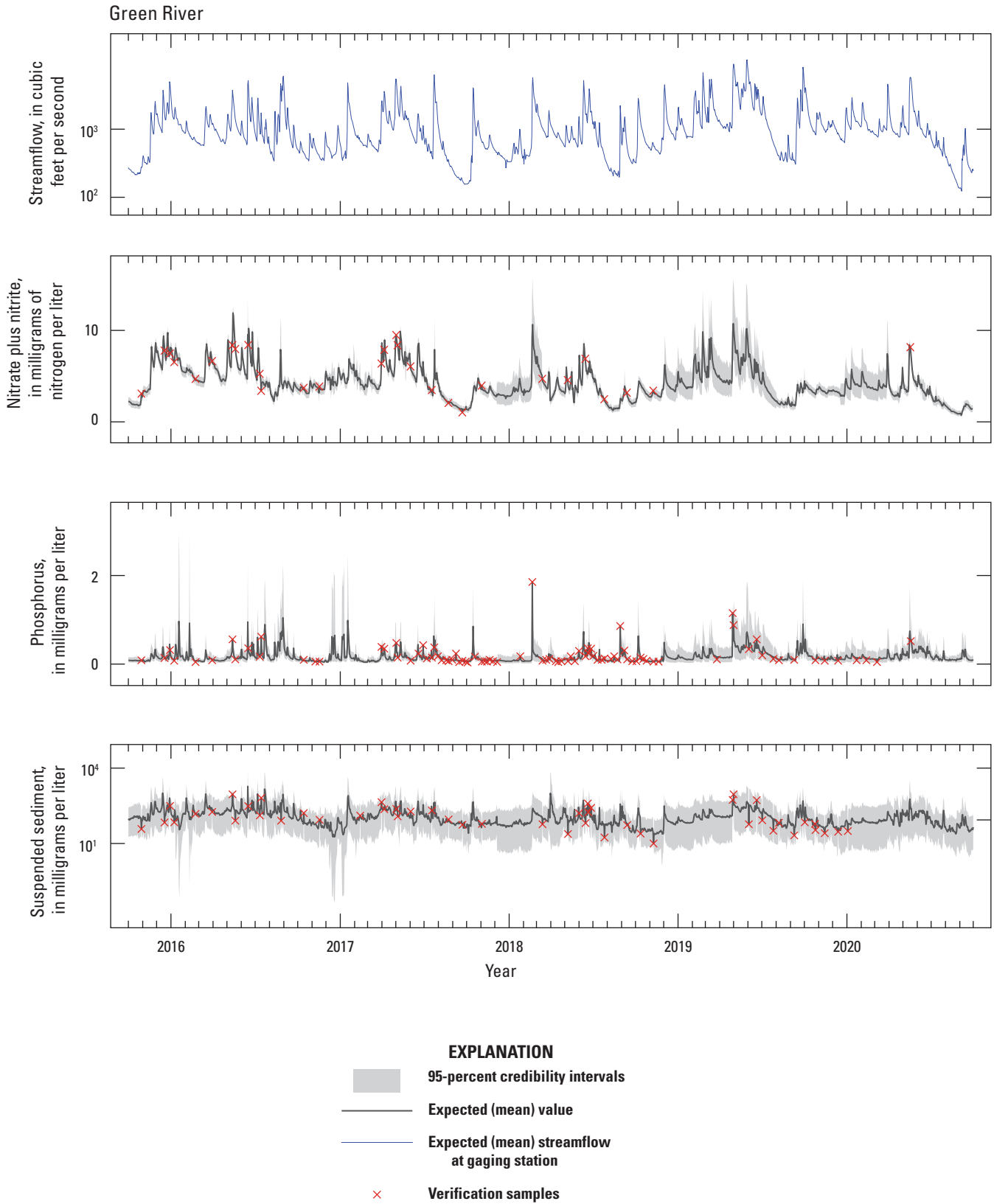


Figure 1.5. Timeseries of streamflow, nitrate plus nitrite, phosphorus, and suspended sediment for the Green River at Geneseo.

Illinois River at Florence, Ill

USGS station number 05586300 (5.4 miles downstream from IEPA site number D-32)

Location—Latitude 39°37'58", Longitude 90°36'28" referenced to North American Datum of 1983, in NE¼SE¼ NW¼ sec. 25, T. 14 N., R. 14 W., Pike County, Ill., hydrologic unit 07130011, on right upstream bridge protection cell of Route 100/106 bridge, 0.3 miles east of Township Road 3982E, 0.3 miles north of Florence, 3 miles southwest of Bloomfield, 62 miles downstream from La Moine River, and mile 56.0 (U.S. Geological Survey, 2021).

Station description—The instrumentation at the site is protected behind a pier protection cell, and the physical configuration allows sensors to be located with adequate depth under all streamflow conditions and during all seasons. Cross-sectional measurements of physical and chemical conditions have demonstrated that the stream channel at this site is well-mixed at all stream stages.

Instrumentation—The station includes the following water-quality monitoring instrumentation:

- YSI EXO 2 with water temperature, specific conductance, dissolved oxygen, pH, and turbidity probes measuring at a 15-minute interval;
- Hach Nitratex nitrate sensor measuring at a 15-minute interval;
- Hach Solitax turbidity sensor measuring at a 15-minute interval;
- Sea-Bird Cycle-PO₄ PO₄ analyzer measuring at a 2-hour interval; and a
- Campbell Scientific CR6 datalogger.

Streamflow characteristics—Streamflow at this station is measured from a nearby gage on the Illinois River at Valley City, Ill., (USGS 05586100) because no major tributaries flow between the two stations. Streamflows can be influenced by upstream streamflow regulation within the Des Plaines River, Chicago Waterway, and Sangamon River, as well as locks and dams on the Illinois River itself. Streamflow required to support commercial shipping is maintained, other than during extreme periods. The station is far enough upstream to prevent backwater influence from the Mississippi River during flood conditions. During the study period, the largest daily mean streamflow was approximately 3,230 m³/s in January 2016 (fig. 1.6), with the largest daily mean streamflow being 3,483 m³/s in May 1943. The lowest daily mean streamflow was approximately 106 m³/s in September 2020, with the lowest recorded being 38 m³/s in September 1984. Streamflow is typically highest in late spring and lowest in late summer and early fall.

Kaskaskia River at New Athens, Ill

USGS station number 05595000 (IEPA site number O-03)

Location—Latitude 38°19'11", longitude 89°53'19" referenced to North American Datum of 1983, in SE¼ NW¼NW¼ sec. 33, T. 2 S., R. 7 W., St. Clair County, Ill., hydrologic unit 07140204, on right downstream side of bridge on battery cell at State Highway 13 at New Athens, 1.4 mi downstream from Silver Creek and at mile 28.4 (U.S. Geological Survey, 2021).

Station description—Instrumentation at the site is well protected, and the physical configuration of the site allows sensors to be located with adequate depth under all streamflow conditions and during all seasons. During the summer months, the channel at the site can become stratified.

Instrumentation—The station includes the following water-quality monitoring instrumentation:

- YSI EXO 2 with water temperature, specific conductance, dissolved oxygen, pH, and turbidity probes measuring at a 15-minute interval;
- Hach Nitratex nitrate sensor measuring at a 15-minute interval;
- Hach Solitax turbidity sensor measuring at a 15-minute interval;
- Sea-Bird Cycle-PO₄ PO₄ analyzer measuring at a 2-hour interval; and a
- Campbell Scientific CR6 datalogger.

Streamflow characteristics—Streamflow at this site is largely dependent on dams and reservoir management upstream, as well as backwater effects from the Mississippi River. Upstream from the site, the Carlyle Reservoir Dam helps to regulate streamflow, and downstream, the Mississippi River can influence streamflows, as well as a lock and dam at the mouth of the Kaskaskia River. Because of these difficult streamflow regimes, index velocity measurements began at the site in 2009. During the study period, the highest mean daily streamflow was about 1,566 m³/s in January 2016 (fig. 1.7). This was the highest streamflow since modern records began in 2009. It is not uncommon for the streamflow to be zero or even slightly negative (reverse streamflow) during low streamflows. However, negative streamflows did not occur during the study. The typical pattern of higher streamflows after spring rain events and low streamflows in the late summer/early fall predominates during the study period and historical record.

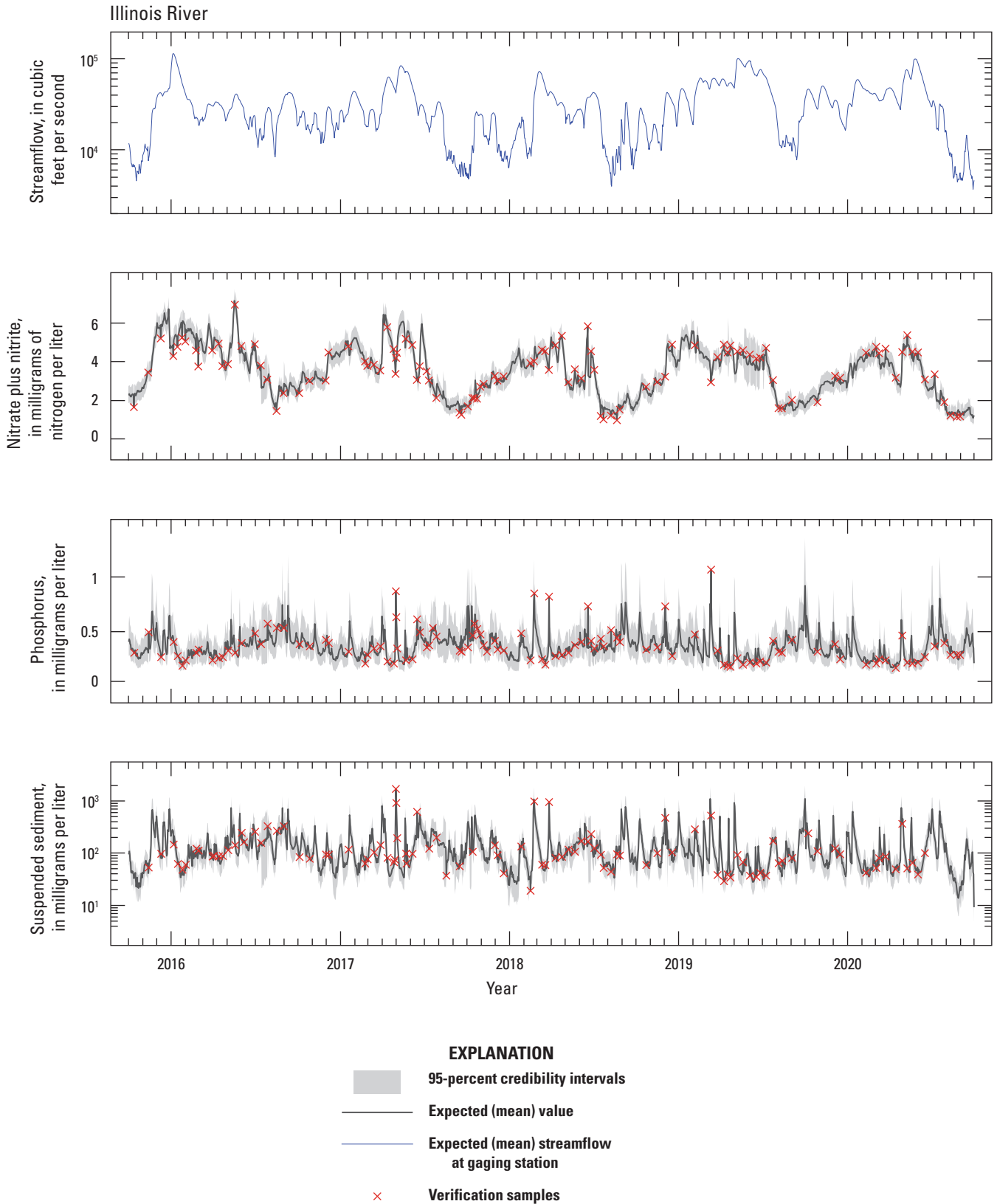


Figure 1.6. Timeseries of streamflow, nitrate plus nitrite, phosphorus, and suspended sediment for the Illinois River at Florence.

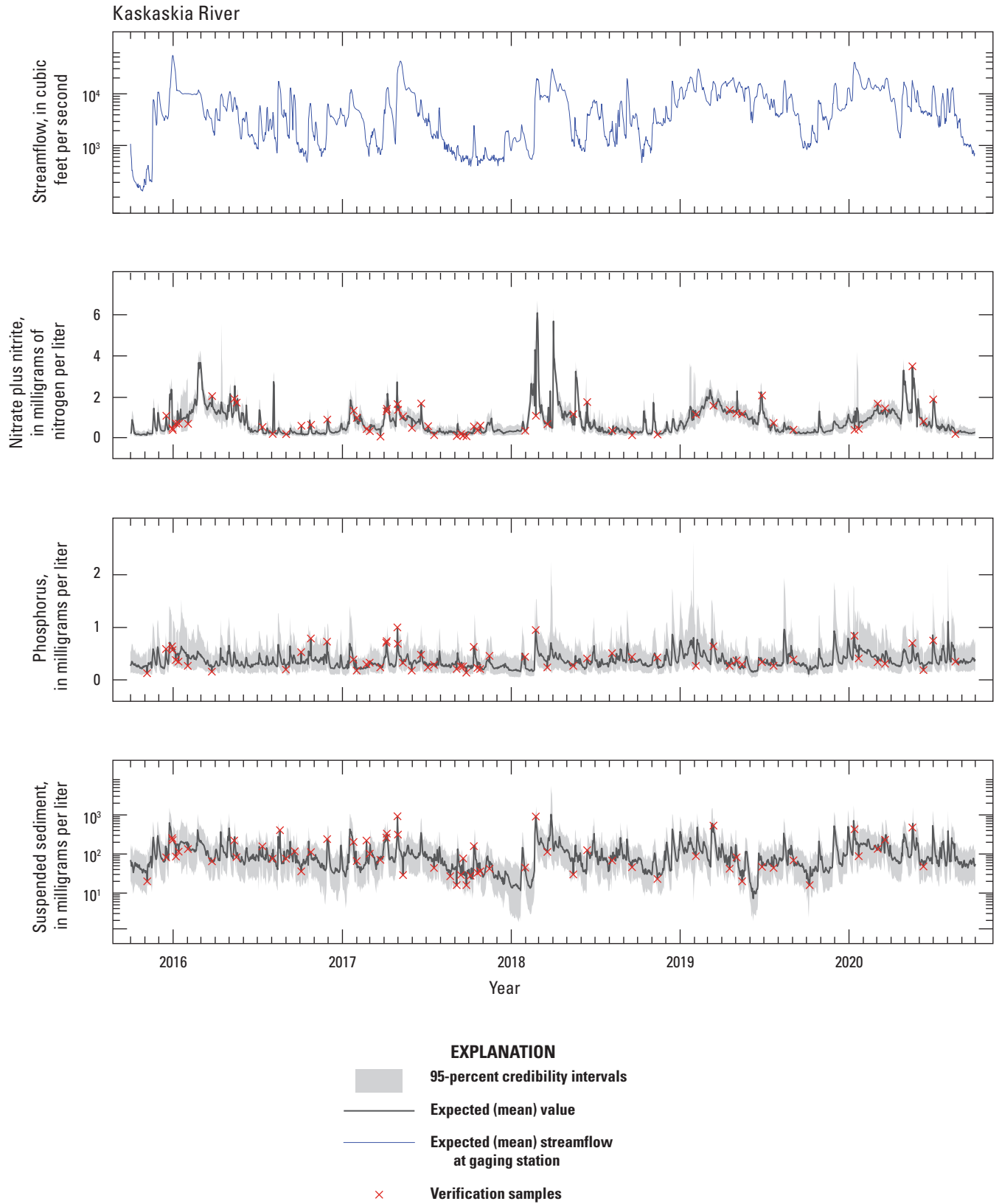


Figure 1.7. Timeseries of streamflow, nitrate plus nitrite, phosphorus, and suspended sediment for the Kaskaskia River at New Athens.

Big Muddy River at Route 127 at Murphysboro, Ill

USGS station number 05599490 (IEPA site number N-12)

Location—Latitude 37°45'30", longitude 89°19'40" referenced to North American Datum of 1983, in NE¼NE¼ sec. 9, T. 9 S., R. 2 W., Jackson County, Ill., hydrologic unit 07140106, on right bank on upstream side of State Highway 127 in Murphysboro, and at mile 37.5 (U.S. Geological Survey, 2021).

Station description—The Big Muddy River station operates reliably with respect to equipment performance and data continuity. One large data gap occurred in the summer of 2017 when sensors were removed during bridge construction. Equipment at the site is well protected, and the physical configuration allows sensors to be located with adequate depth under all streamflow conditions and during all seasons. Water temperatures generally remain warm enough and stream stage fluctuations are suitable for year-round operation. During low summer streamflows, the river can become stratified at this site, but this is unlikely to impact load estimates, which are more reflective of conditions at high streamflows.

Instrumentation—The station includes the following water-quality monitoring instrumentation:

- YSI EXO 2 with water temperature, specific conductance, dissolved oxygen, pH, and turbidity probes measuring at a 15-minute interval;
- Hach Nitratax nitrate sensor measuring at a 15-minute interval;
- Hach Solitax turbidity sensor measuring at a 15-minute interval;
- Sea-Bird Cycle-PO₄ PO₄ analyzer measuring at a 2-hour interval; and a
- Campbell Scientific CR6 datalogger.

Streamflow characteristics—Streamflow at this site is influenced by multiple reservoirs, most notably Rend Lake and Crab Orchard Lake. Reservoir management during storm events, as well as backwater effects from the Mississippi River, can influence streamflow. During the study period, the highest mean daily streamflow was about 818 m³/s in May 2017 (fig. 1.8). Peak streamflow for the site is 1,200 m³/s in May 2011. It is not uncommon for the streamflow to be at zero or negative because of backwater effects of floods downstream on the Mississippi River. During the study period, the largest reverse streamflow event took place in May 2019 at -15 m³/s. The typical pattern of higher streamflows after spring rain events and low streamflows in the late summer/early fall predominates during the study period and historical record.

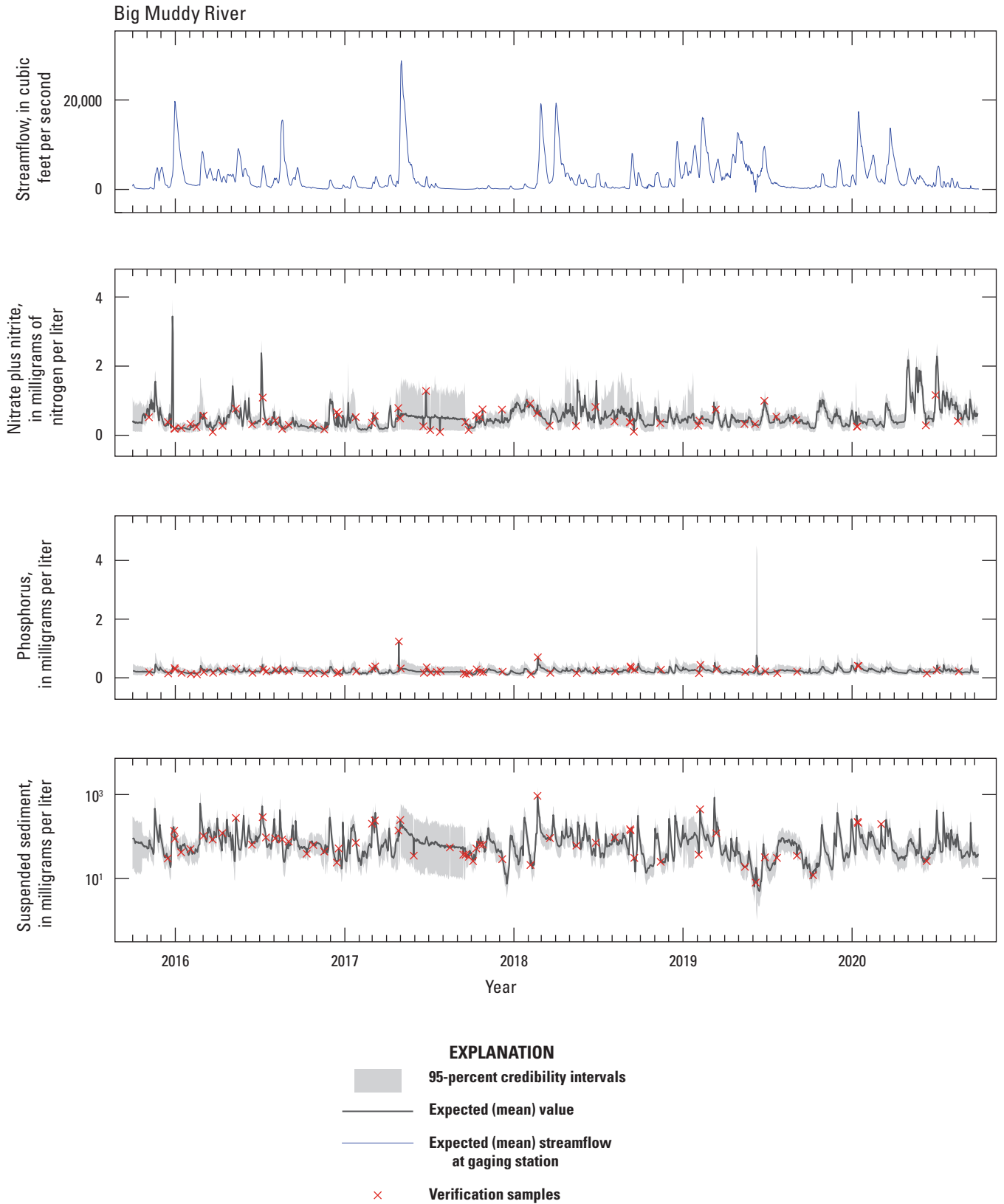


Figure 18. Timeseries of streamflow, nitrate plus nitrite, phosphorus, and suspended sediment for the Big Muddy River at Murphysboro.

References Cited

U.S. Geological Survey, 2021, USGS water data for the Nation: U.S. Geological Survey National Water Information System database, accessed January 2021 at <https://doi.org/10.5066/F7P55KJN>.

For more information about this publication, contact:
Director, [Central Midwest Water Science Center](#)
U.S. Geological Survey
405 N. Goodwin Ave.
Urbana, IL 61801
217-328-8747

Publishing support provided by the Indianapolis and
Lafayette Publishing Service Centers

