JOURNAL OF THE AMERICAN WATER RESOURCES ASSOCIATION

AMERICAN WATER RESOURCES ASSOCIATION

Phosphorus and Nitrogen Transport in the Binational Great Lakes Basin Estimated Using SPARROW Watershed Models

Dale M. Robertson, David A. Saad, Glenn A. Benoy, Ivana Vouk, Gregory E. Schwarz, and Michael T. Laitta

Research Impact Statement: As part of a binational effort, SPARROW watershed models were developed for the entire Great Lakes Basin and used to determine the amount and sources of phosphorus and nitrogen input to each lake.

ABSTRACT: Eutrophication problems in the Great Lakes are caused by excessive nutrient inputs (primarily phosphorus, P, and nitrogen, N) from various sources throughout its basin. In developing protection and restoration plans, it is important to know where and from what sources the nutrients originate. As part of a binational effort, Midcontinent SPARROW (SPAtially Referenced Regression On Watershed attributes) models were developed and used to estimate P and N loading from throughout the entire basin based on nutrient inputs similar to 2002; previous SPARROW models only estimated U.S. contributions. The new models have a higher resolution (~2-km^2 catchments) enabling improved descriptions of where nutrients originate and the sources at various spatial scales. The models were developed using harmonized geospatial datasets describing the stream network, nutrient sources, and environmental characteristics affecting P and N delivery. The models were calibrated using loads from sites estimated with ratio estimator and regression techniques and additional statistical approaches to reduce spatial correlation in the residuals and have all monitoring sites equally influence model development. SPARROW results, along with interlake transfers and direct atmospheric inputs, were used to quantify the entire P and N input to each lake and describe the importance of each nutrient source. Model results can be used to compare loading and yields from various tributaries and jurisdictions.

(KEYWORDS: watershed modeling; loading; nutrients; spatially referenced regression.)

INTRODUCTION

The Laurentian Great Lakes is the largest freshwater system in the world, with nearly 10% of the United States (U.S.) population and 30% of the Canadian population in its watershed (USEPA 2018a). The Great Lakes receive nutrients from many tributaries draining areas ranging from pristine forests to intensively farmed areas and large urban centers, which results in nutrient input from these tributaries being extremely variable (Robertson and Saad 2011). Excessive nutrient inputs have caused eutrophication problems to various degrees and at scales ranging from

Paper No. JAWRA-18-0103-P of the *Journal of the American Water Resources Association* (JAWRA). Received July 23, 2018; accepted July 11, 2019. **Discussions are open until six months from issue publication**.

Upper Midwest Water Science Center (Robertson, Saad), U.S. Geological Survey, Middleton, Wisconsin, USA; Canadian Section (Benoy), International Joint Commission, Ottawa, Ontario, CAN; Ocean, Coastal and River Engineering (Vouk), National Research Council Canada, Ottawa, Ontario, CAN; National Center (Schwarz), U.S. Geological Survey, Reston, Virginia, USA; and U.S. Section (Laitta), International Joint Commission, Washington D.C., USA (Correspondence to Robertson: dzrobert@usgs.gov).

Citation: Robertson, D.M., D.A. Saad, G.A. Benoy, I. Vouk, G.E. Schwarz, and M.T. Laitta. 2019. “Phosphorus and Nitrogen Transport in the Binational Great Lakes Basin Estimated Using SPARROW Watershed Models.” *Journal of the American Water Resources Association* 1–24. https://doi.org/10.1111/1752-1688.12792.

bays around the Great Lakes (e.g., Green Bay in Lake Michigan, Maccoux et al. 2013, and Bay of Quinte in Lake Ontario, Minns et al. 2011) to most of Lake Erie (Watson et al. 2016). Because of the degradation in water quality, several national and binational efforts have been conducted to reduce nutrient loading to the Great Lakes, such as the binational Great Lakes Water Quality Agreement (GLWQA), the Great Lakes Restoration Initiative led by the U.S. Environmental Protection Agency (USEPA 2018b), and the Great Lakes Nutrient Initiative led by Environment and Climate Change Canada (ECCC 2013). As part of the GLWQA (1978), phosphorus (P) loading targets were established for each lake (Table 1). These targets are currently under review as an annex to the GLWQA.

In developing nutrient reduction strategies and restoration plans, it is important to understand where and from what sources the nutrients originate. This information is important to determine the major contributors of nutrients (i.e., identification of hotspots on the landscape) and what types of actions are needed to reduce the loading (i.e., whether to focus on addressing export from point sources, such as wastewater treatment plants [WWTPs], or nonpoint sources, such as agricultural runoff). To describe the spatial variation in P and nitrogen (N) inputs to the Great Lakes and in their sources, Robertson and Saad (2011) developed SPAtially Referenced Regression On Watershed attributes (SPARROW) models (Smith et al. 1997; Schwarz et al. 2006) for the U.S. part of the Great Lakes Basin based on nutrient inputs and landscape practices similar to 2002.

Since publishing the original SPARROW models for the U.S. part of the Great Lakes Basin, there have been several critical evaluations of SPARROW models, model results, and the statistical approaches used to develop the models. One criticism was that the SPARROW models, developed by Robertson and Saad (2011), only described inputs from U.S. watersheds (Richards et al. 2013). Therefore, only U.S. tributaries could be compared and did not provide a complete picture of where nutrients originate in binational waters, such as Lake Erie, and do not enable descriptions of large Canadian watersheds to several lakes, such as the watershed of the Bay of Quinte, Ontario, which experiences eutrophication problems (Minns et al. 2011).

Because the original SPARROW models only provided results for parts of the watershed of four of the Great Lakes, the summary results for these lakes may have provided a biased representation of the relative importance of each of the sources of P and N. This may be especially important in estimating the role of point sources into Lake Erie because the U.S. part of the basin contains most of the point sources (Richards et al. 2013), and inputs from large agricultural areas, such as the Thames River in Canada, were not included.

The stream network used in the original SPARROW models was defined using the enhanced stream-reach file 1 (1:500,000 scale), which resulted in the models having relatively large catchments (median size of ~480 km^2). Therefore, transport from these catchments could not be further subdivided, which makes management decisions that are usually made at scales smaller than this difficult.

Another issue with the original Robertson and Saad (2011) SPARROW models was that they were calibrated with loads estimated using regression techniques, which were later shown to be potentially biased. Regression techniques were shown to often underestimate P loads and possibly overestimate nitrate loads (described below in the Constituent Load Information section) (Stenback et al. 2011; Richards et al. 2013). Therefore, SPARROW models calibrated with these loads could be inaccurate and may provide biased evaluations of the spatial distribution and sources of the loads.

There have also been critical evaluations of the statistical approaches used in calibrating SPARROW models, such as those used to develop the original Great Lakes models. Monitoring is seldom evenly distributed over large study areas, such as the Great Lakes Basin. If all monitored sites are used in model calibration, such as typically used in SPARROW

TABLE 1. Morphometric characteristics, drainage basin size, and total targeted annual phosphorus load for each Great Lake.

Great Lake	Lake area (km^2)[1]	Lake volume (km^3)[1]	Mean depth (m)	Drainage area (km^2)[1]	U.S. drainage area (km^2)	Drainage area-to-surface area ratio	Target phosphorus load[2] (MT/yr — to the entire lake)
Superior	82,100	12,100	147	124,000	43,600	1.5	3,400
Michigan	57,800	4,920	85	116,000	116,000	2.0	5,600
Huron	59,600	3,540	59	132,000	41,400	2.2	2,800
Erie	25,700	484	19	77,500	55,500	3.0	11,000
Ontario	18,960	1,640	86	63,800	35,700	3.4	7,000

Note: U.S., United States.
[1]Encyclopedia Britannica (2019).
[2]Total target loads specified in the Great Lakes Water Quality Agreement of 1978.

calibrations, relationships more characteristic to one area, where site density is high, may drive the overall estimation of the coefficients in the model, and if sites are located too close to one another, measurement error may affect the model calibration. In addition, calibration sites are often nested within the basin of downstream sites (in other words upstream of another monitoring site). When this occurs during typical SPARROW calibration, the model-estimated load at each upstream calibration site is replaced with its monitored load to eliminate errors from propagating down the stream network and to reduce the correlation across the subbasin error terms (Smith et al. 1997). The resulting downstream load that is estimated using the upstream-measured load is referred to as the "conditioned" predicted load used in model calibration, whereas the load completely simulated by the model is referred to as the "unconditioned" predicted load or simply the simulated load. This substitution, however, reduces the magnitude of the errors at the downstream sites, especially when sites have nearby upstream monitored site(s), and can also result in a spatial correlation in the residuals (Qian et al. 2005). This use of conditioned loads reduces the potential influence of the downstream sites on the coefficients in the SPARROW model and can result in an underestimation of the residuals compared to when the model is used to completely simulate loads throughout the basin (Wellen et al. 2014).

There are significant challenges in developing large scale water-quality models, such as SPARROW models that cover the entire Great Lakes Basin, that include multiple jurisdictions and multiple countries. Because of the different geospatial topologies used to describe stream networks and the different enumeration conventions used to describe inputs from the various nutrient sources, harmonization (i.e., the process of bringing together data of varying formats, delineations, naming conventions and transforming them into one cohesive dataset) of each of the datasets used in the model is required for binational modeling. A precursor to the development of SPARROW models for the entire Great Lakes Basin was the development of SPARROW models for the Red-Assiniboine River Basin, a region that includes portions of U.S. and Canada, with three states and two provinces (Benoy et al. 2016). Extensive collaboration between agencies in the U.S. and Canada, coordinated by the International Joint Commission, was necessary to create binational harmonized data layers across the entire stream network. Many of the geospatial and numerical techniques used in the Red-Assiniboine River Basin application of SPARROW are applicable elsewhere in other watersheds that straddle the border.

In this paper, we describe the SPARROW P and N models developed for the entire binational Great Lakes Basin and nearby surrounding areas (collectively referred to as Midcontinent). These models were developed in a manner which eliminates or at least reduces most of the issues with the original Robertson and Saad (2011) SPARROW models mentioned above. Additional statistical approaches to those traditionally used to develop SPARROW models are used to reduce the influence of nonuniformly distributed sites and the effects of sites being nested within the basin of other downstream sites. These SPARROW models can be used to describe nutrient sources and transport in the broader Midcontinental region; however, this paper focuses on the Great Lakes. Here, we describe the P and N inputs from the entire Great Lakes Basin, including previously published estimates of interlake inputs and atmospheric inputs applied directly to the surface of the lakes. This information enables comparisons and ranking of all contributing areas and enables the importance of the sources throughout the entire area of each of the Great Lakes, or any specified smaller area, such as a specific state/province or specific river basin, to be evaluated. Catchments in the new models are based on a much finer stream network enabling transport to be described over much smaller areas than with the original models.

METHODS

Study Area

Midcontinent SPARROW models were developed for the entire binational Great Lakes Basin and nearby Upper Mississippi River, Ohio River, Red River, and Lake of the Woods Basins (Figure 1). Using information from watersheds near, but outside of, the Great Lakes Basin increased the number of calibration sites which increased the range in model input variability and provided a more accurate representation of the full range of conditions within the Great Lakes Basin. Land use and land cover in the study area consists primarily of forests in the northern and southeastern parts and agriculture in the western and central parts. Several major metropolitan areas, including Minneapolis, Minnesota, Chicago, Illinois, Detroit, Michigan, and Cleveland, Ohio, and Toronto, Ontario, are within this area. The Laurentian Great Lakes consist of five lakes linked by relatively short connecting channels. The morphometric characteristics of each lake are given in Table 1. All of the lakes have relatively small drainage area-

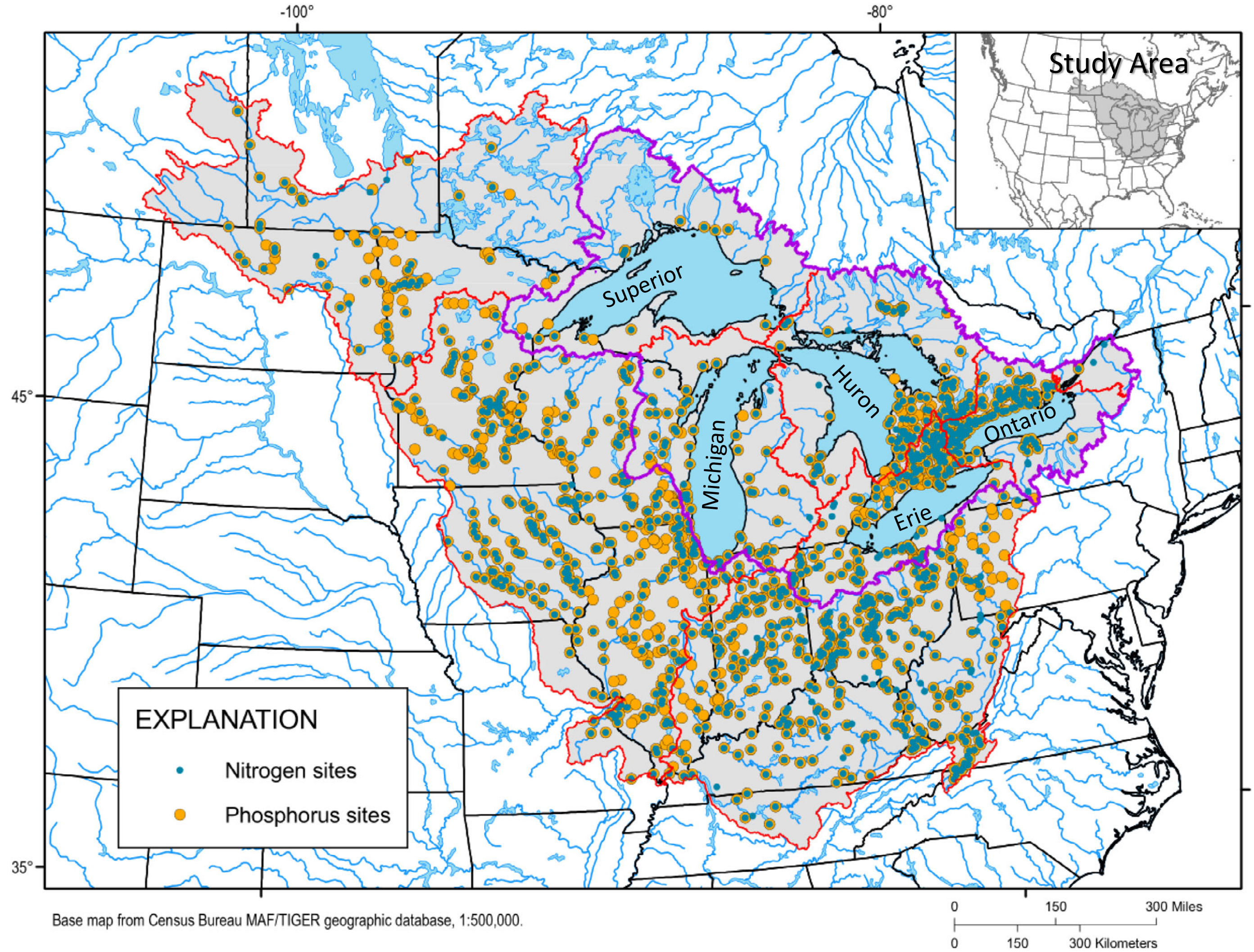

FIGURE 1. Spatial domain of the Midcontinent SPAtially Referenced Regression On Watershed (SPARROW) models, with monitoring sites that were used for model calibration for phosphorus and nitrogen identified.

to-lake surface area ratios, ranging from 1.5 for Superior to 3.4 for Ontario, and have mean depths ranging from 19 m (Erie) to 147 m (Superior).

SPARROW Model

SPARROW is a spatially referenced watershed model that uses a hybrid mass-balance/statistical approach to simulate the nonconservative transport (i.e., includes losses) of a constituent throughout a study area in relation to statistically significant landscape properties, such as climate, soils, and artificial drainage, and instream/reservoir properties (Smith et al. 1997; Schwarz et al. 2006; Alexander et al. 2008). SPARROW models simulate long-term mean-annual transport (loads, the dependent variable in the calibration process) given source inputs and management practices similar to a given base year (in this case 2002). SPARROW models simulate long-term mean-annual constituent transport (i.e., loading) that incorporates a range in hydrological conditions; therefore, they include inputs from both surface-water runoff and groundwater during baseflow and high-flow events. Spatial variability in the environmental setting (described with land-to-water delivery variables) enables variability in the amount of a constituent from each source reaching the stream network. The coefficient reported for each source variable provides an estimate of how much of that source is delivered to streams under the assumption that all spatially variable land-to-water delivery factors are uniformly distributed at average conditions throughout the study area. Part of the constituents that reach the stream is often attenuated or decayed in streams or reservoirs as the constituent travels down the stream network. The amount of a constituent ultimately transported or delivered to a downstream location incorporates the fraction of the inputs delivered to the stream and the fraction delivered during downstream transport, both of which are estimated during calibration.

In SPARROW model development/calibration, a variety of model specifications (regression equations) are evaluated to determine which constituent sources, landscape characteristics, and stream/reservoir decays are statistically significant in controlling constituent transport. In some cases, variables serve as surrogates for other variables that are spatially correlated with the variables specified in the model. For example, although the amount of agricultural land is included as an input source in the N model, it may actually represent N from natural sources, fixation, and other agricultural sources that are not included in the model. Variables identified as statistically significant (typically $p < 0.05$) in explaining the distribution in constituent loads are retained, or if source variables are not statistically significant, they are typically combined with other sources in a series of model calibrations until an acceptable specification is obtained in terms of model fit (root mean square error [RMSE], model-estimated coefficients, variance inflation factors, and residual plots). It is often difficult to completely distinguish the inputs from correlated source variables (such as fertilizer and manure) and this results in large variance inflation factors and relatively large confidence limits on their respective coefficients. Highly correlated important land-to-water delivery variables often results in one of the variables being omitted from the model. Coefficients in the models are typically estimated using nonlinear least squares regression (NLLSR; Schwarz et al. 2006). The least squares methodology used to calibrate the model correctly accounts for the enhanced uncertainty caused by collinear variables (Amemiya 1985; Schwarz et al. 2006). If the estimated coefficients associated with collinear variables are statistically significant in the regression, then the implication is that there was a sufficient number of observations to statistically separate these coefficients from zero.

During calibration, it is optimal to have calibration sites equally distributed throughout the entire study area. Monitoring sites with estimated P and N loads were not evenly distributed over the study area, especially north of Lake Erie (where site density was highest) and north of Lake Superior (where site density was lowest). Therefore, to obtain a more uniform distribution of sites and reduce the effects of measurement errors on closely located sites, calibration sites were thinned to only one site from each 12-digit hydrologic unit code (HUC12; Seaber et al. 1987) in the U.S. or HUC12-equivalent sized areas (~100 km^2) in Canada. The site with the best estimated load, based on its coefficient of variation from the modified Beale ratio estimator (BRE) or regression approaches (Cohn 2005; Lee et al. 2016; discussed in the Constituent Load Information section), was the site selected from each HUC12. This thinning process did not result in a uniform distribution, but it did reduce dense concentrations in some areas (Figure 1).

During calibration, it is optimal for each monitoring site to have similar influence on the determination of coefficients for the variables in the SPARROW model. Because of the way nested sites (a monitoring site located downstream of another monitored site) are typically handled in model calibration, however, the downstream sites tend to have lower residual variance and can result in these sites being underrepresented in the SPARROW statistical calibration process. To address for the potential unequal influence of the nested basins during model calibration, a statistical algorithm was developed in which residual weights are computed as being proportional to the fraction of the upstream drainage area that is downstream of other monitoring sites (the nested area), and these weights are used in a subsequent reestimation of the model using weighted-nonlinear least squares regression, WNLLSR (Schwarz et al. 2006, eq. 1.55). To obtain the weights for each site, the SPARROW model is first calibrated with NLLSR using equal weights applied to all sites to obtain an initial estimate of the model residuals. The squared values of the model residuals are then regressed on the fraction of each monitored basin that is downstream of other monitoring sites (sites with no upstream monitoring sites are assigned a value of 1.0), and if there is a significant relation then the inverse of the predicted values from this regression serve as weights in a subsequent reestimation (recalibration) of the SPARROW model. The specific equation for determining the weights (w_i) is

$$w_i = \frac{N^{-1}\sum_{j=1}^{N}\sigma_j^2}{\sigma_i^2},$$

where N is the number of sites and σ_i^2 and σ_j^2 are the residual variances for the i-th and j-th sites, respectively. The numerator in the weight equation, given by the average of the residual variances, normalizes the weights so that the variance of the residuals from the WNLLSR approximately equals the residual variance in the unweighted regression. Because the coefficient associated with the squared residual and fraction nested area regression relation has a positive sign, this results in a second SPARROW model calibration that uses larger, but more proper, residual weights for the load observations associated with sites that have small areas downstream of other monitoring sites. During this process, the model specification (final variables included in the model and their coefficients) should again be evaluated in an iterative manner. To demonstrate the robustness of the final models, confidence intervals were determined for each coefficient using the standard errors (SEs) from the WNLLSR and the quantile from the standard t distribution.

SPARROW models provide estimations for each stream reach that include incremental (originating in the immediate catchment area) and accumulated (originating in the immediate and all upstream catchments) load and yield, volumetrically weighted concentration, and source-share contributions. In addition, the delivered incremental and accumulated load/yield from any location is described as that part of the load/yield (delivery fraction) ultimately transported downstream to a specific location, in this case each of the Great Lakes, after accounting for downstream removal/attenuation in streams and reservoirs. Loads and yields (with confidence limits) were simulated for all of the Great Lakes, eight-digit HUCs (HUC8s) in the U.S. and sub-subbasins in Canada, and tributaries with drainage areas > 150 km^2 and provided in the Supporting Information to this paper. Because of the nonlinear manner in which the estimated coefficients enter SPARROW models, it was necessary to use bootstrap methods (parametric) to assess uncertainty, which included correcting for potential bias caused by logarithmic retransformations. For a full description of bootstrap methodology, see Schwarz et al. (2006).

Data Used to Calibrate the SPARROW Models

Four types of data are used to "build or calibrate" SPARROW models: stream and reservoir network information to define stream reaches and catchments and to define instream/reservoir decay; long-term mean-annual loads for many sites throughout the study area (dependent variables); information describing all of the main sources of the constituent being modeled (in this case total phosphorus, P, and total nitrogen, N, independent variables); and information describing variability in the environmental characteristics of the study area that causes statistically significant variability in the land-to-water delivery of the constituent (independent variables). All stream and reservoir information and geographic data used to develop the models are described in detail by Vouk et al. (2018a) and constituent loading data are described by Saad et al. (2018) and briefly described below. Great care was taken to harmonize all data across the U.S.–Canada border, and across state/province borders. As a first step in developing approaches to create harmonized binational SPARROW models, a modeling team, including scientists from the U.S. and Canada, developed SPARROW P and N models for the Red-Assiniboine River Basin (Jenkinson and Benoy 2015; Benoy et al. 2016).

Stream Network Information. Water flow paths, incremental reaches, and catchments were defined by streams and water bodies in the 1:100,000-scale National Hydrography Dataset Plus Version 2.0 (Moore and Dewald 2016) for the U.S. and a modified version of the Ontario Integrated Hydrology Data Enhanced Watercourse dataset developed by the Ontario Ministry of Natural Resources and Forestry (2012) for Canada (Vouk et al. 2018a). Additional reservoir information was included from the U.S. Army Corps of Engineers National Inventory of Dams dataset for the U.S. (USACE 2016) and CanVec dataset (Centre for Topographic Information 2012) for Canada. Only reservoirs with surface areas >0.007 km^2 were included. Final datasets included ~820,000 catchments, with a median size of 2.5 km^2 (mean size of 1.2 km^2), with ~265,000 containing reservoirs, compared to ~11,500 catchments, with a median size of ~480 km^2, with only 1,045 containing reservoirs in the original Robertson and Saad (2011) models.

Stream length and mean-annual flow velocity were used to estimate the time of travel used to test for nutrient loss due to natural processes in different-sized streams (categorized as small, medium, and large streams). Phosphorus and Nitrogen removal in various-sized streams were examined in the calibration process, but only medium-sized streams (flow rates of 0.28–2.27 m^3/s; 10–80 cfs) were found to have significant P losses in the final SPARROW P model, and no stream categories had significant N losses in the final SPARROW N model. Reservoir surface area and mean-annual flow were used to compute the inverse of hydraulic loading (flow divided by surface area), which was used to test for P and N removal in reservoirs. Mean-annual flow was obtained from runoff rates reported by McCabe and Wolock (2011) and Wolock and McCabe (2018) for the U.S. and Water Survey of Canada gaging stations data in the Environment and Climate Change Canada (ECCC) HYDAT database for Canada (Vouk et al. 2018a).

Constituent Load Information. Long-term mean-annual loads representing the 2002 base year were computed for sites having adequate streamflow and water-quality data. Load calculation methods and evaluation criteria are described in detail by Saad et al. (2018) and summarized here. These methods and criteria were based on a study by Saad et al. (2011), who evaluated many of the factors affecting the accuracy of load estimation. These methods and criteria attempt to balance the SPARROW model need for numerous sites representing the wide range in watershed characteristics that exist throughout the entire study area with the ability to simulate reasonably accurate loads. Adequate streamflow required sites to have at least 2 years of complete daily flows collected at or near the monitored location between October 1, 1970 and September 30, 2012, including flow data in 2002 (the base year for the model). Data were compiled from the U.S. Geological Survey (USGS) National Water Information System database (U.S. Geological Survey 2017) for the U.S. and ECCC HYDAT database (ECCC 2016) for Canada. An expansive effort was made to inventory, evaluate, and compile water-quality data from numerous federal, state/provincial, tribe, regional government agencies, and nongovernmental organizations in both U.S. and Canada. Adequate water-quality data required a site to have at least 25 P or N samples during 1970–2012, which minimally overlapped with streamflow data representing that site for 2 years. All water-quality records must be within a predefined proximity to 2002, which depended on the length of the water-quality record. See Saad et al. (2018) for details on the temporal proximity requirements.

For each site, long-term mean-annual P and/or N loads were computed using the Fluxmaster program (Schwarz et al. 2006), which estimates loads using both a regression approach (Cohn 2005) and modified BRE approach (Lee et al. 2016). Both approaches take advantage of measured water quality collected over a range of flow conditions to estimate water quality for days when the site was not monitored, including days with high flows when much of the nutrients is transported. The BRE approach is typically used to compute annual loads; however, in Fluxmaster, the BRE approach was used to estimate the long-term mean-annual load using eight strata formed by subdividing daily average flows from all years into two classes (delineated by the 80th percentile of flow) and four seasons. The regression-based approach used a five-variable water-quality model that was a function of flow, seasonality (sine and cosine terms), trend, and an intercept. Both the BRE load and regression-based load detrended to 2002 were computed for each site.

The original SPARROW models (Robertson and Saad 2011) were calibrated with loads computed using only the regression approach, which was later shown to occasionally provide inaccurate and possibly biased loads (Stenback et al. 2011; Richards et al. 2013; Hirsch 2014). To minimize this problem, long-term mean-annual BRE-computed loads computed using the entire period of flow were used whenever there was no trend in the loads because this approach was shown to have little bias and better at estimating long-term mean-annual loads than most regression approaches (Lee et al. 2016). For sites with no trend in loads, the BRE-computed loads were on average 18% higher than the regression-computed P loads and 8% lower for N loads, which is consistent with that found by Richards et al. (2013). If there was a significant ($p < 0.05$) trend in load, a mean-annual load detrended to 2002 computed using the regression approach was considered for use in model calibration. Prior to use, all loads were evaluated for accuracy and bias. Load relations with SE > 50% of the mean load estimate were considered unacceptable, which is consistent with the accuracy level used in previous SPARROW studies (Saad et al. 2011), and dropped from consideration. Potential biases in regression-computed loads were calculated as the ratio (R) of BRE load divided by the regression load. For loads having trends, R can deviate from 1.0 even if load estimates have no bias because of the trend in water quality and the proximity of the center of the water-quality record compared to the base year. Therefore, a trend-adjusted acceptable bias range was computed using a trend factor m, where $m = \exp$ (trend in load × number of years from the midyear of the dataset), in years away from 2002. In staying consistent with a 50% acceptable difference in the SE of the loads, the acceptable range for the ratio was $m/1.5 < R < m \times 1.5$. If R was outside this range, the regression-computed load was considered unacceptable, and dropped.

Although, much nutrient monitoring has been conducted, only a small subset of these sites had sufficient data to compute loads using our criteria. After evaluation of 35,000 sites with nutrient data, there were 1,425 sites for which P loads were computed and 1,289 sites for which N loads were computed. Therefore, if the goal of monitoring was to describe nutrient loading, more comprehensive monitoring protocols were needed. Approximately 70% of the loads were computed using the BRE approach. Most of the final load sites easily exceeded the minimum flow and water-quality criteria selection protocols. More than 95% of the sites had at least 10 years of flow record. More than 50% of the P loads had at least 146 water-quality samples covering 25 years or more; 80% had at least 60 samples covering more than 11 years. More than 50% of the N loads had at least 125 water-quality samples covering 26 years or more; 80% had at least 51 samples covering more than 12 years. While there were no criteria explicitly used to identify sites that represent the range in flow and seasonal conditions, the final set of load sites also represented these conditions reasonably well. The ratio of average flow on sampled days to average flow for the full flow record was used to evaluate how well the range in flow was represented in the monitored loads. Ratios > 1 indicate that samples, on average, represent higher flow conditions and ratios < 1 indicate samples represent lower flow conditions. Average and median ratios were close to 1 (the average was 0.99 for P and 1.04 for N and medians for both were 1.0). Seasonal conditions were evaluated by looking at the number of sites with at least three samples in each of the four seasons: 86% of the P sites and 87% of N sites had at least three samples collected in each of the four seasons. Based on these evaluations, the loads were believed to represent the range in flow and seasonal conditions reasonably well.

There was wider range in watershed sizes for sites with loads used in the current model calibration than used in the original models: the 5th percentile was ~45 km^2 and 95th percentile was ~36,500 km^2 compared to a 5th percentile of ~156 km^2 and 95th percentile of ~33,200 km^2 for sites used by Robertson and Saad (2011).

Phosphorus and Nitrogen Source Information. Input to the SPARROW models included data that attempt to describe or quantify all the major sources of P and N. After evaluating a variety of model specifications, the final P model included six P sources: WWTPs, urban and open/barren areas (collectively referred to as urban areas), farm fertilizers, manure, agricultural land, and forest/wetland (collectively referred to as forested areas). The area of agricultural land was used to represent all nonfertilizer and nonmanure sources in agricultural areas, such as natural sources and increased erosion because of agricultural activities. The final N model included five N sources: WWTPs, urban areas, farm fertilizers, manure, and atmospheric deposition. Since there is no general agricultural term in the N model, all increased losses of N in agricultural areas are included in the fertilizer and manure terms. Since there is no atmospheric deposition in the P model, this source would be captured in the other sources. Both models also had direct input from the Missouri River in the U.S. and the Qu'Appelle River in Canada, which contributed P and N from outside the study area. Their inputs were estimated from gaged locations at the study-area boundary. Phosphorus or Nitrogen inputs as point sources, land applied, or as related to land use characteristics for each catchment were estimated for the 2002 base year, or as close to that year as possible. All sources and specified land use characteristics are described in detail by Vouk et al. (2018a) and described briefly here.

WWTP effluent data (flow and P and N concentration data) were obtained from the U.S. Environmental Protection Agency Permit Compliance System database supplemented with additional data obtained directly from the states of Wisconsin and Minnesota for the U.S. and Ontario Clean Water Agency and Ministry of the Environment and Climate Change for Ontario. Phosphorus and Nitrogen effluent loads were then computed for each WWTP with or without concentration data (concentrations for sites without measurements were estimated based on typical pollutant concentrations for the magnitude of flow for the facility) using methods described by McMahon et al. (2007), Hoos et al. (2008), and Vouk et al. (2018a). Manitoba and Saskatchewan effluent loads were calculated based on population in the catchment and an export per capita rate based on data obtained from Manitoba Conservation and Water Stewardship, and Saskatchewan Ministry of Environment and the Saskatchewan Water Security Agency (Vouk et al. 2018a).

Fertilizer and manure inputs to each catchment were estimated from total county applications in the U.S. or sub-subbasins in Canada and the amount of agricultural land in each county or sub-subbasin. Inorganic farm fertilizer (referred to as fertilizer) inputs in the U.S. were based on 2002 county-level estimates from Ruddy et al. (2006) and in Canada were based on 2001 Census of Agriculture (AAFC 2002, 2007) data by census division and census consolidated subdivision. Manure inputs were estimated from 2002 county livestock head counts from the Census of Agriculture (NASS 2004; Mueller and Gronberg 2013) for the U.S., and by sub-subbasin estimates for 2001 from the Interpolated Census of Agriculture (AAFC 2001) for Canada. Phosphorus and Nitrogen manure coefficients for each animal type were obtained from Ruddy et al. (2006).

Land use-related inputs (export from a general land type) were based on the amount of the catchment (in km^2) in each general land type (i.e., urban, agriculture, and forested) represented in the 2001 National Land Cover Data (Homer et al. 2007) for the U.S. and the Natural Resources Canada Geobase land cover-data circa 2002 (Centre for Topographic Information 2009) for Canada. Export coefficients for each land type were estimated in the SPARROW calibration process.

Total atmospheric deposition of N in 2002 onto each catchment was estimated from Congestion Mitigation and Air Quality Improvement (CMAQ) Program deposition rates (Schwede et al. 2009; Hong et al. 2011; USEPA 2018b) and the area of the catchment.

Environmental Setting Variables. Statistical methods, similar to those used for determining which P and N sources to include in the models, were used to identify which characteristics were important in explaining variability in P and N delivery to streams. Many characteristics thought to affect nutrient delivery to streams were examined in determining which statistically significant land-to-water delivery factors to include in the models. Environmental setting data were summarized into an average value for each catchment (e.g., average air temperature) or percentage of the catchment with a specific characteristic (e.g., percent clay content).

Environmental setting variables found to significantly influence land-to-water delivery of P were air temperature and percent soil clay content. For the N model, significant variables included air temperature, catchment runoff, and percent of the catchment underlain with tile drains. Mean air temperatures, representing the 1971–2000 average, were obtained from the PRISM database (PRISM Climate Group 2009) for the U.S. and from the Canadian Forest Service (McKenney et al. 2006) for Canada. The average percent clay in the soil was estimated from data in the U.S. Department of Agriculture STATSGO database using methods described by Wolock (1997) for the U.S. and estimated from data from Soil Landscape of Canada version 2.2 (AAFC 1996) for Canada; areas of missing data in Canada were estimated with assistance from the Agriculture and Agri-Food Canada. Mean-annual runoff for 1971–2000 was estimated from annual runoff values from the Wolock and McCabe (1999, 2018) water balance model for the U.S. This model was also used to estimate mean-annual runoff from 1971 to 2000 from Canadian catchments using air temperature from the Canadian Forest Service (McKenney et al. 2006) and precipitation from the Canadian Precipitation Analysis (National High Impact Weather Laboratory 2014). Percent of the catchment with tile drains in the U.S. was based on early 1990s tile drain information compiled by Nakagaki et al. (2016) following the methods described by Sugg (2007). Percent of the catchment with tile drains in Ontario catchments was obtained from Ontario Ministry of Agriculture Food and Rural Affairs (2015) data, and in Manitoba catchments from Sustainable Development (Vouk et al. 2018a). Because of the aridity of Saskatchewan, tile drains were assumed to be absent in this province. All environmental setting data used in the models are described in detail by Vouk et al. (2018a).

Atmospheric Deposition on the Lakes and Interlake Transport. Total atmospheric deposition of N onto each lake was computed from their surface areas and total deposition rates for 2002 from the CMAQ Program (Schwede et al. 2009; Hong et al. 2011; USEPA 2018b). Total deposition of P onto each lake for 2002 was computed by linearly detrending annual depositions to each lake published by Dolan and Chapra (2012) to 2002. Linearly detrending annual data to 2002 was done by regressing the annual deposition rates on years and determining the deposition value for 2002. Contributions from the adjacent upstream lake(s) were estimated from annual interlake transport estimated by Dolan and Chapra (2012), who estimated interlake transfer of P from measured flow between the lakes and observed in lake concentrations. Interlake P input in 2002 was estimated by linearly detrending the annual inputs. Interlake N inputs were estimated from the interlake P input and the ratio of total N to total P measured during spring in each lake from Dove and Chapra (2015) for the 2–3 years centered on 2002. Total N was based only on the sum of nitrite plus nitrate and ammonia. Organic N was not included because it was not measured during routine monitoring; therefore, the interlake transfer of N is biased low.

RESULTS

Calibration of SPARROW Models

Preliminary SPARROW P and N models were first developed using all load sites that met the defined accuracy criteria (the typical approach used in developing SPARROW models), which identified an initial list of potential source, land-to-water delivery, and instream decay variables. The initial variables included in this traditional model-development approach ("all data") are given in Table 2, which provides an example of the full calibration process for the P model. Next, the sites were thinned down to one site per HUC12-sized area, and the variables included in the model were reevaluated. This thinning process reduced the number of load sites from 1,425 to 1,197 for P and from 1,289 to 1,101 for N (the final, thinned sites used for model development are shown in Figure 1). This step dropped basin slope from the model, and it affected the magnitude of a few coefficients for the variables, especially the WWTP coefficients ("thinned" in Table 2). For the final calibration step, the model residuals were examined to see if they were significantly related to the fraction of the basin that was nested. The residuals were significantly related to the fraction of the basin that was nested ($p < 1.0\ e^{-23}$); therefore, the weight of each thinned-calibration site was adjusted based on the fraction of its drainage area that was downstream of other monitoring sites and the model was reevaluated. During typical SPARROW calibrations, all sites are equally weighted. The weights applied to the sites in nested-weighting calibration ranged by about a factor of 10 for P and a factor of 5 for N, with sites having small areas downstream of other monitoring sites having larger weights. Model specification

TABLE 2. Summary of model development for the phosphorus SPARROW model.

Variable/Summary statistic	Variable units	Model coefficients		
		All data	Thinned data set	Thinned and adjusted for nesting (final model[1])
Sources				
Wastewater treatment plants (WWTPs)	kg	1.152	1.020	1.155
Urban and open areas	km^2	82.71	90.01	82.08
Fertilizers (farm)	kg	0.018	0.019	0.020
Manure	kg	0.022	0.026	0.026
Agricultural sources (other)	km^2	40.39	28.16	23.21
Forest, wetland, and shrubland	km^2	17.52	17.72	18.20
Land-to-water delivery				
Air temperature	°C	0.083	0.083	0.087
Clay content	%	0.019	0.019	0.018
Basin slope	degrees	0.049	NI	NI
Aquatic loss				
Instream decay (0.3–2.3 m^3/s)	m^3/s	0.262	0.328	0.243
Reservoir loss	yr/m	5.104	4.867	5.685
Summary statistics				
Unconditioned RMSE		0.654[2]	0.654	0.650[3]
Number of sites		1,425	1,197	1,197

Note: Calibration incorporated adjustments for the high density of sites in specific areas (thinned) and the amount of the upstream watershed downstream of other calibration sites (nested area) (NI, not statistically significant in model; therefore, not included in the model).
[1]Full final model results are provided in Table 3.
[2]Root mean square error (RMSE) computed only for the thinned 1,197 sites.
[3]Not adjusted for different weighting applied to the sites.

in the last two steps was done in an iterative manner, until the final variables were chosen.

The final SPARROW P model was a function of: six P sources (WWTPs, urban land, farm fertilizers, manure, agricultural land, and forested land); two land-to-water delivery factors (air temperature and percent clay content); nutrient removal in medium-sized streams; and deposition in reservoirs (Tables 2 and 3). The coefficients for all sources and factors were highly significant ($p < 0.02$), indicating that each source, land-to-water delivery factor, and instream/reservoir factor was important in describing the distribution in measured loads. The coefficients were also robust (see the 90% confidence intervals for the coefficients in Table 3). This P model had a conditioned RMSE of 0.603 (based on comparisons of measured and conditioned predicted loads in natural logarithmic units). Based on full model predictions (based on unconditioned simulated loads), the unconditioned RMSE was 0.624. The distribution of the unconditioned residuals is shown in Figure 2a. A few regional biases in model unconditioned simulated loads occurred but were mostly outside of the Great Lakes Basin (overestimations in the extreme the northwest and southeast and underestimations in the southwest); however, areas north of the Great Lakes, where few calibration sites exist, were also underestimated.

In SPARROW models, the export of each source from a catchment, except point sources, is usually modified by land-to-water delivery factors. The values of the coefficients associated with each source (Table 3) describe the average effect of the land-to-water delivery factors, but spatial variability in these factors enable variability in transport of the sources to streams across the study area. WWTP effluent is directly added to the stream network. The coefficient for WWTP (1.155) being >1.0 suggests that either WWTP input may have been underestimated, or more likely that this variable also incorporates the effects of other sources, such as commercial and industrial point sources and combined sewer overflows. The calibrated model indicates that urban areas on average contribute ~82 kg/km^2/yr in addition to contributions from WWTPs. There were three agricultural sources, namely fertilizers, manure, and a general agricultural area source. The fertilizer and manure coefficients suggest that ~2.0% of the input from farm fertilizers reach streams compared to ~2.6% for manure (coefficients = 0.020 and 0.026, respectively), although these percentages were not statistically different. The calibrated model indicates that agricultural areas contribute an additional ~23 kg/km^2/yr to that from fertilizers and manure, which represents general losses from agricultural areas, such as natural sources and increased losses caused by agricultural activity. On average, forest, wetland, and shrubland areas are estimated to contribute ~18 kg/km^2/yr. The P model did not include atmospheric deposition; therefore, this source would be included in the other defined sources.

TABLE 3. Summary of SPARROW model calibration results.

Variable/Summary statistic	Variable units	Coefficient units	Model coefficient value	90% Confidence interval for the model coefficicients low	90% Confidence interval for the model coefficicients high	Standard error of the model coefficient	Probability level (p value)	Variance inflation factors
Phosphorus model								
Sources								
WWTPs	kg	fraction, dimensionless	1.155	0.830	1.480	0.198	<0.0001	1.2
Urban and open areas	km^2	kg/km^2/yr	82.1	64.7	99.5	10.6	<0.0001	1.8
Fertilizers (farm)	kg	fraction, dimensionless	0.020	0.010	0.030	0.006	0.0005	8.8
Manure	kg	fraction, dimensionless	0.026	0.019	0.033	0.004	<0.0001	2.3
Agricultural sources (other)	km^2	kg/km^2/yr	23.2	9.00	37.4	8.64	0.0037	8.8
Forest, wetland, and shrubland	km^2	kg/km^2/yr	18.2	15.2	21.2	1.82	<0.0001	1.6
Land-to-water delivery								
Air temperature	°C	°C^{-1}	0.087	0.065	0.110	0.014	<0.0001	2.5
Clay content	%	dimensionless	0.018	0.012	0.024	0.004	<0.0001	1.9
Aquatic loss								
Instream decay (0.3–2.3 m^3/s)	m^3/s	days^{-1}	0.243	0.055	0.431	0.114	0.0168	1.7
Reservoir loss	yr/m	m/yr	5.685	4.324	7.045	0.827	<0.0001	1.3
Summary statistics								
Weighted RMSE		0.603						
Unconditioned weighted RMSE		0.624						
Mean exponentiated weighted error		1.197						
Number of sites (1,479 original sites)		1,197						
Nitrogen model								
Sources								
WWTPs	kg	fraction, dimensionless	0.540	0.401	0.678	0.084	<0.0001	0.6
Urban and open areas	km^2	kg/km^2/yr	617	464	769	92.7	<0.0001	0.8
Fertilizers (farm)	kg	fraction, dimensionless	0.108	0.089	0.127	0.011	<0.0001	2.7
Manure	kg	fraction, dimensionless	0.115	0.093	0.137	0.013	<0.0001	1.1
Atmospheric deposition	kg	fraction, dimensionless	0.193	0.169	0.218	0.015	<0.0001	2.0
Land-to-water delivery								
Runoff	Ln(mm/yr)	dimensionless	0.652	0.576	0.728	0.046	<0.0001	2.8
Air temperature	°C	°C^{-1}	−0.051	−0.065	−0.036	0.009	<0.0001	2.5
Tile drains — percent of catchment	%	dimensionless	0.012	0.011	0.014	0.001	<0.0001	2.5
Aquatic Loss								
Reservoir loss	yr/m	m/yr	1.210	0.578	1.843	0.385	0.0017	1.9
Summary statistics								
Weighted RMSE		0.390						
Unconditioned weighted RMSE		0.425						
Mean exponentiated weighted error		1.088						
Number of sites (1,289 original sites)		1,101						

Note: Calibration incorporated adjustments for the amount of the upstream watershed that was included in upstream calibration sites.

Reckhow et al. (1980) and Beaulac and Reckhow (1982) conducted a literature search of nutrient export rates for selected land uses and found P export of 2–82 kg/km^2 (median of 22 kg/km^2) from forested areas, 19–623 kg/km^2 (median of 108 kg/km^2) from urban areas, and 10–1,860 kg/km^2 (medians of 225 kg/km^2 from row crops and 100 kg/km^2 from pasture) from agricultural areas. Our estimated yields

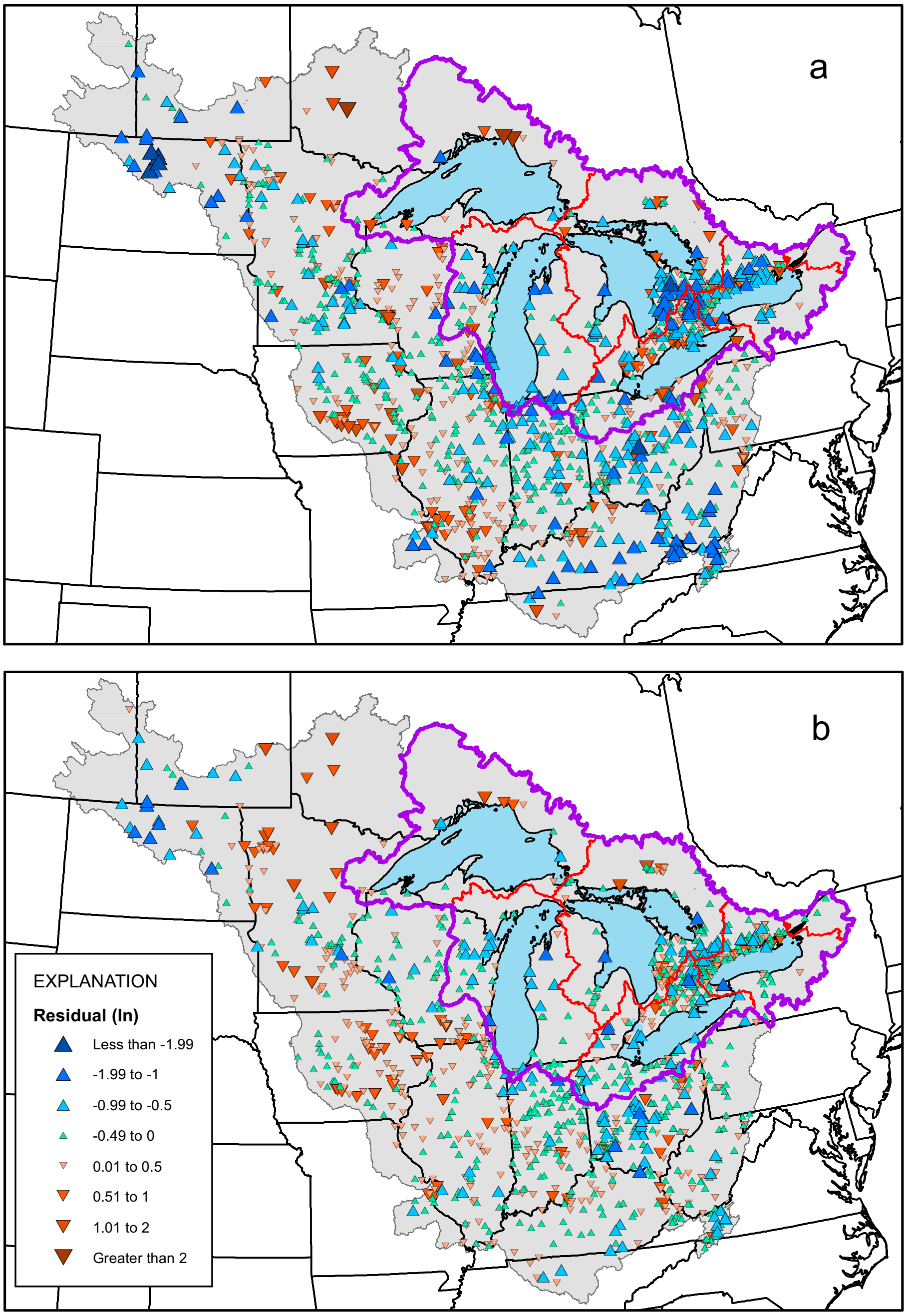

FIGURE 2. Predictability of the Midcontinent (a) phosphorus and (b) nitrogen SPARROW models. All residuals are in natural logarithmic units. All major basins are delineated.

for forested, urban, and agricultural areas are in the range of those published. Our export rates from urban areas did not include WWTP effluent and from agricultural areas do not include export from fertilizers and manure, which likely resulted in our estimated rates being on the low end of those published.

Based on the sign of the land-to-water delivery coefficients (positive values reflect enhanced delivery), P yields were higher in areas with higher air temperatures and higher clay content. Streams were subdivided into three sizes based on their average with mean-annual streamflow ($<0.3\ m^3/s$; $0.3–2.3\ m^3/s$; and $>2.3\ m^3/s$). Instream losses were only significant in medium-sized streams ($0.3–2.3\ m^3/s$; $8–80\ ft^3/s$), and not significant in small and large streams. The optimum range of streams with instream losses was found by iteratively changing the ranges to minimize p values during model calibration. P removal in reservoirs was also significant.

A similar approach to that used for P was used to develop the final SPARROW N model. The final N model had: five N sources (WWTPs, urban areas, farm fertilizers, manure, and atmospheric deposition); three land-to-water delivery factors (runoff, air temperature, and percent of the catchment underlain by tile drains); and one factor describing removal in reservoirs (Table 3). Nitrogen from general agricultural activities was not found to be significant; therefore, it would be included in the fertilizer and manure terms. N losses in different-sized streams were examined but were insignificant. The residuals were significantly related to the fraction of the basin that was nested ($p < 1.4\ e^{-12}$); therefore, the weight of each thinned-calibration site was adjusted based on the fraction of its drainage area that was downstream of other monitoring sites and the model was reevaluated. A significant coefficient describing N losses in medium-sized streams was initially found when all monitoring sites were included, but this coefficient was insignificant when the load data were adjusted for upstream nesting; therefore, instream decay was not included in the final model (discussed later). Coefficients for all variables in the final N model were highly significant ($p < 0.002$) indicating that each source, land-to-water delivery factor, and reservoir factor was important in describing the distribution in measured N loads. The coefficients were robust (see the 90% confidence intervals in Table 3). This model had a conditioned RMSE of 0.390, and an unconditioned RMSE of 0.425. The distribution of unconditioned residuals is shown in Figure 2b. The only consistent regional biases in unconditioned loads were mostly outside of the Great Lakes Basin (overestimated in the far northwest).

Source coefficients in the N model indicate that ~54% of the estimated N input from WWTPs, ~11% of farm fertilizers, ~12% of manure, and ~20% of the atmospheric deposition reach the stream network (Table 3). Relatively more of the agricultural N sources (11%–12%) were transported to streams than the agricultural P sources (2%–3%). In addition to effluent from WWTPs, urban areas contribute ~616 kg/km^2/yr. The WWTP coefficient (0.54) being <1.0 suggests that our WWTP input may have been overestimated. This is not too surprising given that most N inputs from WWTPs were based on typical pollutant concentrations. Additional inputs from agricultural and forested lands were not significant in the N model. Inputs from fixation and other agricultural losses are likely to be included with other sources that they may be correlated, such as fertilizers and manure. Losses from forested areas are likely to be included with the atmospheric deposition source. Based on the signs of the land-to-water delivery coefficients, N yields should be higher in areas with higher runoff, cooler air temperatures, and more tile drains. Removal/deposition in reservoirs was also significant in the model.

Delivered Incremental Nutrient Yields

Delivered incremental P and N yields (load per unit catchment area delivered to a specified downstream target, such as the Great Lakes) from each catchment are shown in Figure 3. Delivered incremental yields are mediated by the amount and type of nutrients input to the catchment and by land-to-water delivery, stream, and reservoir factors affecting their transport. Highest delivered yields were from catchments with WWTPs, such as Detroit, Michigan, and Chicago, Illinois; however, lower but still relatively high-delivered P yields (>88 kg/km^2/yr) and N yields (>1,140 kg/km^2/yr) were from catchments in extensive agricultural areas. Within the Great Lakes Basin, highest P and N yields were primarily around Erie and the southeast shore of Huron and lowest north of Superior. The main differences in geographic patterns of P and N yields can be explained by differences in agricultural source distributions, with higher P yields from areas dominated by animal agriculture. Accumulated P and N loads and yields representing the 2002 base year, with confidence intervals, for each Great Lake, HUC8/sub-subbasin, and tributaries > 150 km^2 are provided in the Supporting Information. It should be noted that these loads/yields were not adjusted for the prediction errors shown in Figure 2, that is, they are the nonconditioned loads and yields.

Total Phosphorus and Nitrogen Delivery to the Great Lakes

To reduce the effects of prediction errors from the SPARROW models (Figure 2) when estimating the total loads delivered to each Great Lake, model results were used only to simulate loads from unmonitored areas and measured loads were used wherever they were available. This combination of measured and modeled loads is referred to as the "total conditioned" load. The total conditioned P loading from the watershed ranged from 1,610 MT/yr (metric tonnes per year) into Superior to 8,900 MT/yr into Erie (Table 4 and Figure 4a). The percent of the total conditioned load obtained from monitored sites varied among lakes depending on the number and location of the monitoring sites. Contributions of P from Canada ranged from 46%–47% for Superior, Huron, and Ontario, to 23% for Erie, to 0% for Michigan. Full SPARROW-simulated P loads for each lake ranged from being 25% lower to 36% higher than the conditioned loads. The total conditioned N loading ranged from 24,500 MT/yr (Superior) to 155,000 MT/yr (Erie; Table 4 and Figure 4b).

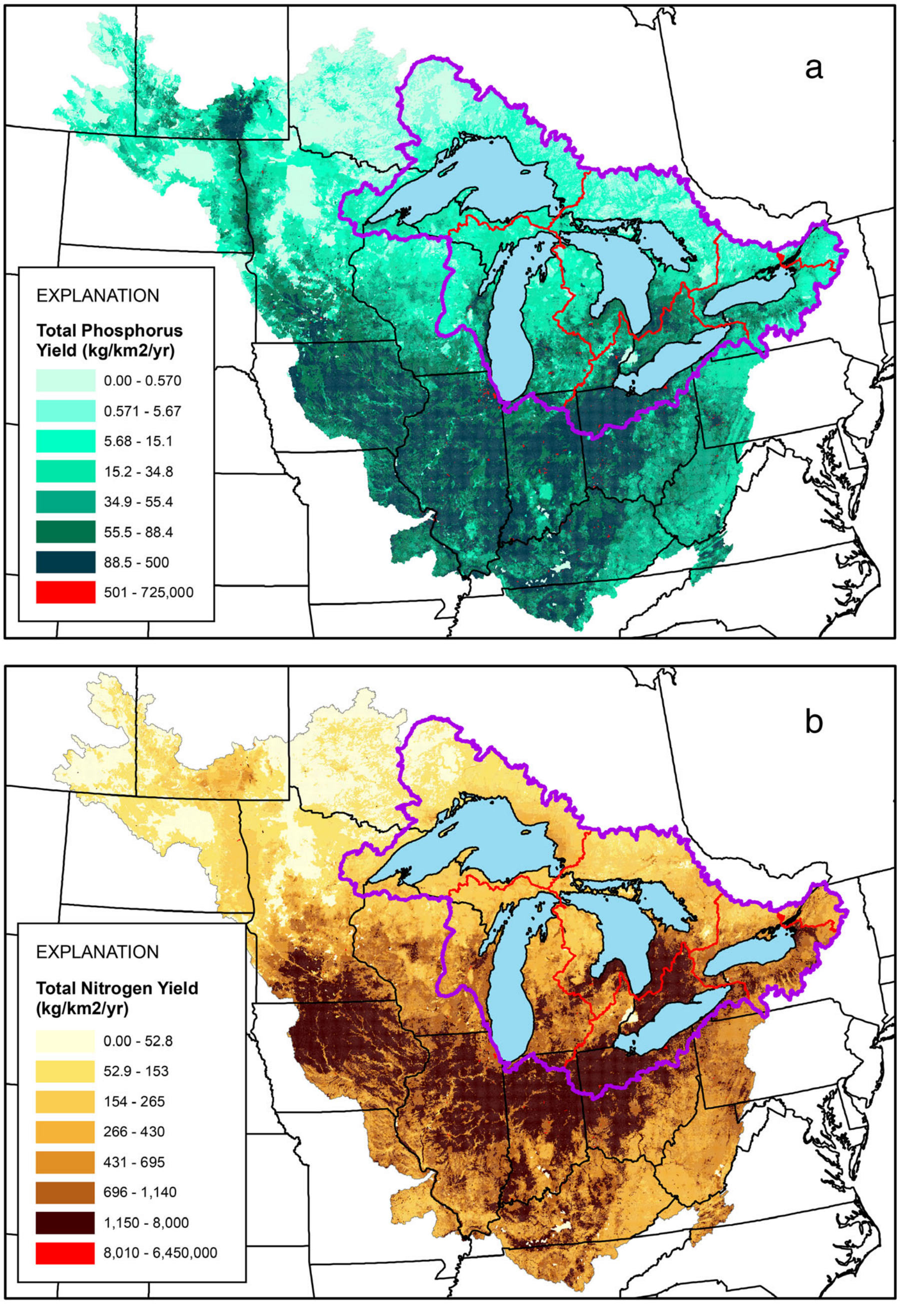

FIGURE 3. Delivered incremental yields (full model predictions) of (a) phosphorus and (b) nitrogen, in kg/km^2/yr.

Contributions of N from Canada ranged from 56–63% for Superior and Huron, to 31%–40% for Erie and Ontario, to 0% for Michigan. Full SPARROW-simulated N loads to each lake ranged from being 4% lower to 23% higher than the total conditioned loads.

Annual P yields from the watershed ranged from 11.1 kg/km^2 (Superior) to 114 kg/km^2 (Erie) and annual N yields ranged from 169 kg/km^2 (Superior) to 1,990 kg/km^2 (Erie; Table 4). Phosphorus and Nitrogen yields from the Erie watershed were much higher than from other lake watersheds, which coincide with it having the highest percentage of agriculture in its watershed. Yields were second highest from the Ontario watershed, followed by Michigan, Huron, and Superior.

After including P and N from nontributary sources (interlake transfer and direct atmospheric deposition on the lakes), total annual P loading was 2,420 MT to Superior, 4,080 MT to Michigan, 3,140 MT to Huron, 9,860 MT to Erie, and 4,940 MT to Ontario (Figure 4 and Table 5). All loads are below the targets established by the GLWQA (Table 1). After including N

TABLE 4. Total annual watershed load (conditioned and fully simulated with the model), in MT to each Great Lakes, total and by country.

Lake	Total area: Conditioned load (MT)	Total area: Full model simulated load (MT)	Total area: Percent difference in loads	Total area: Conditioned yield (kg/km^2)	U.S.: Conditioned load (MT)	U.S.: Percent of conditioned total	U.S.: Full model simulated load (MT)	U.S.: Percent difference in loads	Canada: Conditioned load (MT)	Canada: Percent of conditioned total	Canada: Full model simulated load (MT)	Canada: Percent difference in loads
Phosphorus load												
Superior	1,610	1,210	−24.6	11.1	860	53.4	757	−12.0	751	46.6	457	39.2
Michigan	3,770	4,850	28.6	32.4	3,770	100.0	4,850	28.6	0	0.0	0	0.0
Huron	2,320	3,150	35.6	17.3	1,230	52.7	1,650	34.3	1,100	47.3	1,510	37.0
Erie	8,900	9,030	1.5	114.0	6,890	77.4	7,320	6.3	2,010	22.6	1,710	−15.1
Ontario	3,340	3,180	−4.6	52.9	1,800	54.1	1,610	−10.9	1,530	45.9	1,580	2.8
Nitrogen load												
Superior	24,500	23,500	−3.8	169	10,600	43.4	10,300	−2.7	13,900	56.6	13,200	−4.6
Michigan	67,600	83,000	22.8	581	67,600	100.0	83,000	22.8	0	0.0	0	0.0
Huron	75,000	80,400	7.2	559	27,600	36.8	32,100	16.3	47,400	63.2	48,300	1.8
Erie	155,000	173,000	11.9	1,990	107,000	69.0	126,000	17.8	48,000	31.0	47,400	−1.1
Ontario	54,800	59,500	8.6	868	33,000	60.3	34,000	2.9	21,700	39.7	25,500	17.2

Note: Conditioned loads are estimated with measured loads and simulated loads from SPARROW for ungaged areas.

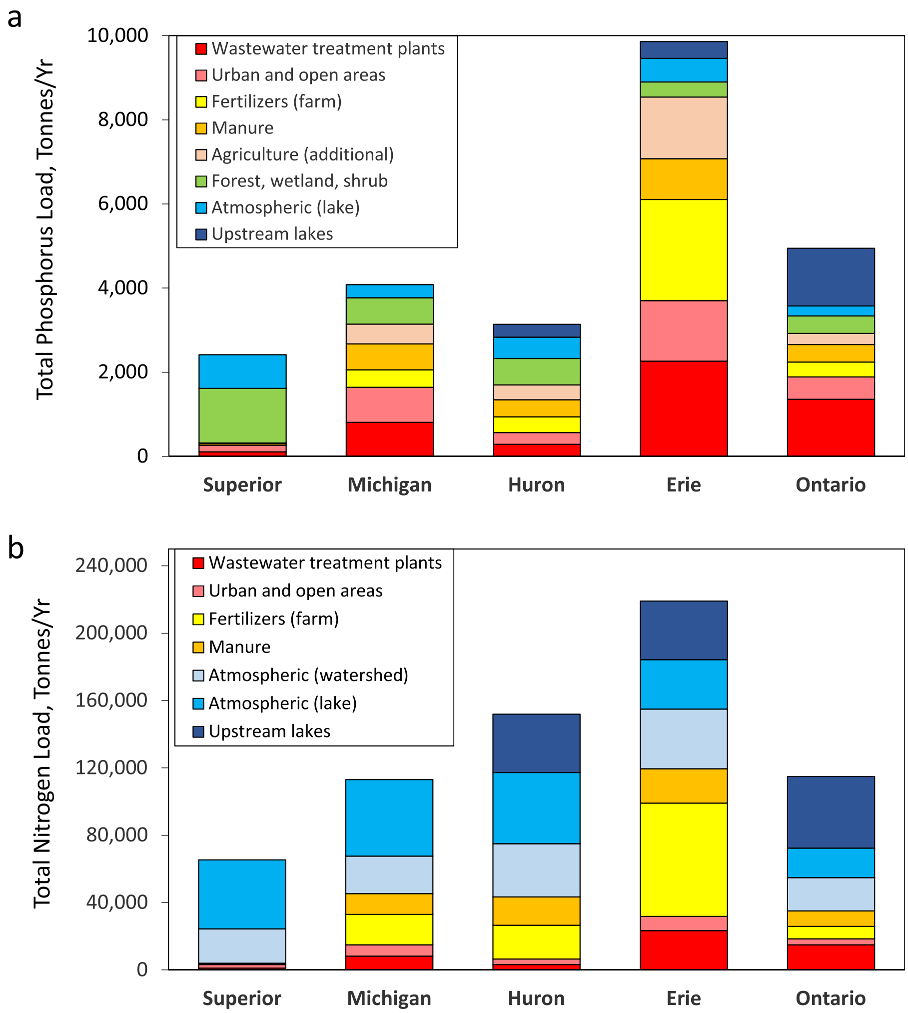

FIGURE 4. Delivered incremental loads (based on conditioned watershed loading) of (a) phosphorus and (b) nitrogen, in metric tonnes (MT)/yr, subdivided by inputs by source.

from nontributary sources, total annual N loading was 65,300 MT to Superior, 113,000 MT to Michigan, 152,000 MT to Huron, 219,000 MT to Erie, and 115,000 MT to Ontario.

Sources of Phosphorus and Nitrogen to the Great Lakes

Sources of P in the SPARROW model included inputs from WWTPs, urban land, farm fertilizers, manure, other agricultural sources, and forested land. The percentage of the total P load delivered to each lake originating from each source is shown in Figure 4a and given in Table 5, along with a breakdown into broad land use categories: urban, agriculture, and forest sources. Each lake had a different ranking in the relative importance of individual P sources. The largest broad land use source was forest sources for Superior, urban sources for Michigan and Ontario, and agricultural sources for Huron and Erie. The relative importance of each agricultural source is shown for each lake in Figure 4. In general, manure was the largest agricultural P source for all lakes, except Erie that was dominated by fertilizers (Table 5).

The importance of P from WWTPs and other urban sources varied from 16%–24% (Superior and Huron), 42%–44% (Michigan and Erie), to 57% (Ontario) of the watershed loading (Table 5; Figure 4a). This is a decrease in importance compared to that estimated by Robertson and Saad (2011) (discussed below). Nontributary loadings (interlake transfer and direct atmospheric deposition) were important for some lakes. Over 30% of the total P loading to Superior was from direct atmospheric deposition, and over 27% of the P entering Ontario was from Erie.

Sources of N in the SPARROW model included inputs from WWTPs, urban land, farm fertilizers, manure, and atmospheric deposition (Table 5; Figure 4b).

TABLE 5. Annual watershed loading (conditioned) and complete lake loading of phosphorus and nitrogen to each of the Great Lakes, and percent by source.

				Percentage by specific source									Percentage by broad landuse category		
Lake	Total loading (MT)	Watershed area (km^2)	Yield (kg/km^2)	WWTP	Urban land	Fertilizers (farm)	Manure	Agricultural land	Forest land	Atmosphere (watershed)	Atmosphere (direct on lake)	Upstream lake	Urban sources	Agricultural sources	Atmosphere/forest/upstream lake
Phosphorus loading															
Watershed loading															
Superior	1,610	145,000	11.1	6.6	9.8	0.3	2.2	0.7	80.5	NI	0.0	0.0	16.4	3.2	80.5
Michigan	3,770	116,000	32.4	21.4	22.0	11.1	16.2	12.5	16.6	NI	0.0	0.0	43.5	39.9	16.6
Huron	2,320	134,000	17.3	12.2	12.1	16.0	17.6	15.2	26.9	NI	0.0	0.0	24.3	48.8	26.9
Erie	8,900	78,200	113.8	25.4	16.2	27.0	10.9	16.4	4.1	NI	0.0	0.0	41.6	54.3	4.1
Ontario	3,340	62,800	53.2	40.6	16.1	10.5	12.5	7.9	12.5	NI	0.0	0.0	56.6	30.9	12.5
Complete lake loading															
Superior	2,420	82,100	29.4	4.4	6.5	0.2	1.4	0.5	53.7	NI	33.3	0.0	10.9	2.1	87.0
Michigan	4,080	57,800	70.6	19.8	20.4	10.3	15.0	11.6	15.3	NI	7.6	0.0	40.1	36.9	23.0
Huron	3,140	59,600	52.6	9.1	8.9	11.9	13.0	11.2	19.9	NI	16.1	9.8	18.0	36.1	45.9
Erie	9,860	25,700	383.6	22.9	14.6	24.4	9.9	14.8	3.7	NI	5.6	4.1	37.5	49.1	13.4
Ontario	4,940	19,000	260.7	27.4	10.8	7.1	8.4	5.3	8.4	NI	4.9	27.6	38.2	20.8	40.9
Nitrogen loading															
Watershed loading															
Superior	24,500	145,000	169	4.6	8.3	0.8	2.5	NI	NI	83.8	0.0	0.0	12.9	3.3	83.8
Michigan	67,600	116,000	581	12.1	10.1	26.7	18.2	NI	NI	33.0	0.0	0.0	22.1	44.9	33.0
Huron	75,000	134,000	559	4.3	4.3	26.7	22.6	NI	NI	42.2	0.0	0.0	8.6	49.2	42.2
Erie	155,000	78,200	1,980	15.1	5.4	43.5	13.1	NI	NI	22.9	0.0	0.0	20.5	56.7	22.9
Ontario	54,800	62,800	872	27.2	6.6	13.5	16.8	NI	NI	36.0	0.0	0.0	33.7	30.3	36.0
Complete lake loading															
Superior	65,300	145,000	451	1.7	3.1	0.3	0.9	NI	NI	31.4	62.5	0.0	4.8	1.2	93.9
Michigan	113,000	116,000	972	7.2	6.0	16.0	10.9	NI	NI	19.7	40.2	0.0	13.2	26.8	59.9
Huron	152,000	134,000	1,133	2.1	2.1	13.2	11.1	NI	NI	20.8	27.8	22.8	4.2	24.3	71.5
Erie	219,000	78,200	2,801	10.7	3.8	30.8	9.3	NI	NI	16.2	13.5	15.9	14.5	40.0	45.5
Ontario	115,000	62,800	1,830	12.9	3.1	6.4	8.0	NI	NI	17.1	15.3	37.0	16.1	14.4	69.5

Note: NI, not included in the model.

Agricultural sources were the dominant broad source of N for Michigan, Huron, and Erie, atmospheric deposition on the land was the dominant source for Superior, and there was no dominant broad source for Ontario. In general, fertilizers were a more important source of N than of P. Farm fertilizers were the largest agricultural source of N for Michigan, Huron, and Erie, whereas manure was the largest agricultural source for Ontario. It should be noted that the SPARROW N model did not directly distinguish export from natural sources, fixation, and other agricultural sources. Urban sources of N were most important for Ontario and least important for Huron. Urban sources of N were relatively less important than for P and only represented 9%–34% of the watershed loading.

Nontributary N loadings were very important for all lakes. Direct atmospheric deposition represented between 13%–15% (Erie and Ontario) and 62.5% (Superior) of total N loading. Over 35% of the N entering Ontario was from Erie. Nitrogen contributions from interlake transfers were more important than indicated here; their inputs were underestimated because organic forms of N were not included in routine monitoring.

DISCUSSION

Before SPARROW models that traverse international borders could be created, multi-jurisdictional (e.g., binational) datasets had to be assembled. In North America, the U.S. and Canada have, to varying extents, developed their own monitoring and research programs, which have resulted in different hydrological, water quality, and geospatial datasets. For example, different sets of water-quality constituents are sometimes measured by different federal, state, or provincial agencies and measured at different frequencies. Similarly, census data reporting units and spatial resolution often differ between the U.S. and Canada. Consequently, a critical preparatory phase of this binational SPARROW modeling project was the establishment of an international team of water quality, geospatial, and modeling experts, who, as part of a precursor study, developed protocols to harmonize many different types and sources of datasets for the binational Red-Assiniboine River Basin (Jenkinson and Benoy 2015; Benoy et al. 2016). This rather unique binational team then adapted these protocols to assemble data and develop SPARROW models for the entire Midcontinental area of Canada and the U.S. as depicted in Figure 1, which enabled the entire international P and N input to each of the Great Lakes to be quantified and the importance of each nutrient source to be described.

Comparisons of Results with Previous Studies

The original SPARROW models developed by Robertson and Saad (2011) for the Great Lakes only described loading from the U.S. part of the watershed. With Midcontinent SPARROW models, which include the entire Great Lakes Basin and portions of the provinces of Saskatchewan, Manitoba, and Ontario, total P and N loading was estimated for each lake (Tables 4–6). To reduce the effects of prediction errors in the models (Figure 2) in estimating the total P and N load to each lake, SPARROW results were only used to simulate loads from unmonitored catchments, and measured loads were used wherever they were available (conditioned loads). The conditioned watershed loadings are compared with those estimated by Dolan and Chapra (2012), after their watershed loads were linearly detrended to 2002, in Table 6. Overall, the loads compare very well (within 16%), except for Superior that was 55% less than estimated by Dolan and Chapra (2012). The similarity in total load estimates primarily may have been the result of both studies using the BRE technique to compute long-term mean-annual loads. The larger difference in loads estimated for Superior is partially due to limited monitoring sites around Superior and uncertainty in the SPARROW model in this area (Figure 2).

Conditioned SPARROW P and N loads from only the U.S. part of the basin are compared with conditioned SPARROW loads estimated by Robertson and Saad (2011) in Table 6. Phosphorus loading to the lakes with the current model increased 0.2%–33.1% from the previous estimates. The largest increase in loading occurred for Erie, which had the largest percentage of its watershed monitored. Much of the increase was due to changes in the estimated loads for the monitored tributaries. In general, P loads computed for this study, using BRE techniques, increased from those computed with regression techniques (discussed below). This increase in P loads using BRE is consistent with that found by Lee et al. (2016), and was primarily the result of regression techniques fitting a linear equation in logarithmic space to a nonlinear flow-to-concentration relation. Changes in N loads between models were less consistent, with decreases in loading for three lakes (Superior, Michigan, and Erie; −2.7% to −27.2%) and increases in two lakes (Ontario and Huron; 0.7% to 6.3%). Inconsistent differences in the changes in N loading between regression and BRE techniques are also consistent with that found by Lee et al. (2016).

Total tributary P loading to Ontario was estimated for 1976, 1989, 1993, and 2008 by Makarewicz et al. (2012), using various methods for monitored and unmonitored areas. Differences in flows and methods

TABLE 6. Comparison of total and U.S. watershed loading (conditioned) of phosphorus, P, and nitrogen, N, to each of the Great Lakes from this study with those of Dolan and Chapra (2012) and Robertson and Saad (2011).

	Entire Great Lake watershed — P			U.S. watershed — P					U.S. watershed — N				
	SPARROW (this study)	Dolan and Chapra (2012)		SPARROW (this study)		Robertson and Saad (2011) SPARROW			SPARROW (this study)		Robertson and Saad (2011) SPARROW		
Lake	Load (MT/yr)	Load (MT/yr)	Percent difference in loads	Load (MT/yr)	Percent nonurban sources	Load (MT/yr)	Percent nonurban sources	Percent difference in loads	Load (MT/yr)	Percent nonurban sources	Load (MT/yr)	Percent nonurban sources	Percent difference in loads
Superior	1,610	2,500	55.1	860	77.4	782	76.1	9.1	10,600	8,400	10,900	9,200	−2.7
Michigan	3,770	3,180	−15.7	3,770	56.5	3,430	53.2	9.0	67,600	2,600	70,000	54,300	−3.6
Huron	2,320	2,190	−5.8	1,230	62.1	927	46.2	24.3	27,600	23,200	25,900	22,500	6.3
Erie	8,900	8,460	−4.9	6,890	51.0	4,610	47.9	33.1	107,000	79,000	136,000	102,000	−27.2
Ontario	3,340	3,190	−4.4	1,800	54.3	1,800	49.5	0.2	33,000	26,600	32,800	21,600	0.7

among years resulted in total annual tributary loading ranging from 2,052 to 6,155 MT, with an average of 3,598 MT. Their average annual loading is close to the 3,340 MT (Table 6) estimated in this study. Their most recent total P load to Ontario for 2008 (2,606 MT) is less than that estimated in this study, but the relative contributions from Canada and the U.S. are similar.

The structures of the SPARROW models developed in this study were much different than those developed for the binational Red-Assiniboine River Basin models. The larger area of the Midcontinent models allowed additional monitored sites and spatial variability that enabled more complete SPARROW models to be developed, with more source and land-to-water delivery variables. Coefficients of the source variables that were in common in these SPARROW models were not expected to be the same because the average delivery of nutrients to streams, demonstrated by the model coefficients, in these two areas were expected to be different. The additional sites also enabled several coefficients, which were included in the Red-Assiniboine Basin models as constants (e.g., WWTP and atmospheric deposition), to be estimated.

Instream attenuation of both P and N was found to be important in previous SPARROW models developed for this area (Robertson and Saad 2011, 2013); however, these models were based on stream and reservoir datasets that were much less detailed than those used in this study. With much more detailed stream and reservoir information used in the models developed in this study, many of the catchments contained lakes and reservoirs with a wider range in size. This resulted in much of the nutrient attenuation being allocated to reservoirs rather than to the streams and rivers.

Change in Load Estimation Techniques and Its Effects on Source Allocations

About 70% of the calibration sites used in this study had loads computed using the BRE technique (Saad et al. 2018). Many BRE-computed P loads increased from those previously estimated using regression techniques. Two sites used in the Robertson and Saad (2011) SPARROW models and shown by Richards et al. (2013) to have P loads greatly underestimated compared to those computed by the National Center for Water Quality Research (NCWQR) using extensive daily data were the Maumee and Sandusky Rivers. Annual loads computed by NCWQR were detrended to 2002 for this comparison (Table 7). For both sites, regression-estimated loads used in calibration of the original SPARROW models were much less than those computed by NCWQR (underestimated by 14%–29% for loads computed using regression techniques and used in SPARROW calibration and were underestimated in the SPARROW-simulated loads by 44%–47%). However, loads used in current model calibration and simulated with the current model were much closer to NCWQR loads (overestimated by 5%–15% in loads computed using BRE and used in calibration and were overestimated by 2%–12% in SPARROW-simulated loads). Therefore, using BRE techniques to compute loads for sites without trends in concentrations appear to improve monitored load estimation and the accuracy of the SPARROW models.

Richards et al. (2013) stated that underestimating P loads used to calibrate SPARROW models should lead to an overestimation in the relative importance of point sources. To evaluate this possibility, the importance of WWTPs to Maumee and Sandusky River loads was examined (Table 7). Based on the Robertson and Saad (2011) SPARROW model, WWTPs contributed 18.8% and 9.3% of these loads, respectively, compared to 10.0% and 5.2% when NCWQR loads were used (assuming 100% of WWTP input was delivered to Erie). In the current model, WWTPs were estimated to contribute less of the total load, 7.9% and 4.2%, respectively, compared to 8.8% and 4.3% when the NCWQR loads were used (note that the percentages for NCWQR changed because the WWTP estimates were modified). The importance of urban sources of P also became less important in the current models (Table 6). Therefore, it appears that the conclusion of Richards et al. (2013) was correct. The current approach using BRE-computed loads to calibrate SPARROW models appears to have eliminated much of this potential bias in loads and the importance of point source inputs.

Including inputs from Canada, interlake transfers, and direct atmospheric deposition to the total P and N budgets also decreased the relative importance of urban inputs to the Great Lakes. Most of these additional inputs were from nonurban sources. Therefore, these complete P and N budgets provide a much better representation of the relative importance of each source to the lakes. Data from these SPARROW models also provide information to examine the relative importance of each source at a range of spatial scales (such as Great Lake, HUC8-sized basin, and tributaries to the Great Lakes > 150 km^2, which are provided in the Supporting Information).

Effects of Additional Model Calibration Statistics

There have been critical evaluations of the approaches typically used to calibrate SPARROW models, mainly that using a nonuniform distribution of sites and sites nested within the basin of other downstream sites can result in certain areas and sites having an unequal influence on the final models (Qian et al. 2005; Wellen et al. 2014). To reduce this unequal influence of the sites, preliminary models were first developed using all load sites (typical calibration approach). The coefficients of this preliminary P SPARROW model are shown in Table 2, and results from this preliminary model are shown by lake in Figure 5 (first bars for each lake). Then, to minimize the effects of areas with a higher density of calibration sites and minimize the effects of measurement errors, the calibration sites were thinned down to one site per HUC12-sized area, and the variables included in the model were reevaluated. This step had little effect on which variables were statistically significant in the models, overall RMSE (based only on sites in both

TABLE 7. Comparison of long-term average phosphorus loads at two sites in the Lake Erie Basin computed by the National Center for Water Quality Research (NCWQR), computed in this study, and simulated with SPARROW, and the resulting change in the relative importance of inputs from WWTPs.

	NCWQR		SPARROW calibration data		SPARROW-simulated results			
Site	Load (MT)	WWTP percent of load	Computed load (MT/yr)	Percent difference from NCWQR load	Modeled load (MT/yr)	Percent difference from NCWQR load	WWTP load (MT/yr)	WWTP percent of SPARROW-simulated load
Comparison with Robertson and Saad (2011) SPARROW model results								
Maumee River	2,090	10.0	1,480	−29.2	1,110	−47.0	209	18.8
Sandusky River	420	5.2	360	−14.3	234	−44.3	22	9.3
Comparison with SPARROW model results from this study								
Maumee River	2,090	8.8	2,200	5.2	2,340	11.8	185	7.9
Sandusky River	420	4.3	485	15.4	427	1.7	18	4.2

Note: U.S. Geological Survey station numbers: Maumee River at Waterville, Ohio, 04103500 and Sandusky River near Fremont, Ohio, 04198000.

models), and total loads to each lake, but it did affect the values of a few coefficients resulting in a small change in the relative importance of the various sources (second bars in Figure 5). Next, weights were applied to the residuals from each site based on the fraction of its drainage area downstream of other monitoring sites, and the models were reevaluated (third bars in Figure 5). This step also had little effect on the overall RMSE and the total load to each lake, but it did cause a few land-to-water variables and an instream decay variable in the N model to become insignificant, and therefore these variables were dropped from the final models. This step again resulted in the source coefficients changing and the importance of each source to also change. By thinning the number of sites in areas with a dense monitoring network and adjusting the weights of sites because of other upstream monitoring sites, we feel that the final SPARROW models better represent the entire study area and especially instream and reservoir losses. Although a few variables were dropped from the models with these additional evaluation steps, we feel that this approach results in each of the monitoring sites, especially the sites with other nearby upstream monitoring sites, having a more similar influence in developing the final model, which is what the statistics behind the SPARROW model assumed. The variables that were dropped in the final models may have initially been included because the residuals of sites with nearby upstream monitoring sites were estimated to be unrealistically small. These additional steps in model calibration did not dramatically affect the overall results of Midcontinental SPARROW models (Figure 5), possibly due to the large number of sites included in the calibration of these models. However, this approach could have a more dramatic effect on models developed with fewer calibration sites, a more nonhomogenous distribution of sites, or developed in areas where most monitoring sites occur along major rivers. These additional steps in calibration should be incorporated into developing future SPARROW models. Alternative Bayesian approaches have also been developed to account for nonuniform site distributions and sites that are nested within the basins of other downstream sites (Qian et al. 2005; Wellen et al. 2014; Alexander 2015).

Adjusting for Potential Model Errors

All spatial models have a certain degree of regional spatial biases and local model errors (differences in measured and modeled loads) in their predictions. To reduce the effects of local errors in the SPARROW models (Figure 2) when estimating total P and N loads delivered from the watershed to each Great Lake, SPARROW results were used only to simulate loads from unmonitored catchments and measured loads were used wherever they were available (referred to as total conditioned loading). When computing the source fractions for the total conditioned loading, the original source fractions for the monitored sites were unchanged, but because the loads for these sites increase or decrease, source fractions of the final total conditioned loads change slightly. Total conditioned P loading to each lake is given in Table 4 and represented in Figure 5 (fourth bars). The conditioned P loads ranged from being 25% more than that fully modeled (Superior) to 36% less than that fully modeled (Huron). The total conditioned N loads ranged from being 4% more than that fully modeled (Superior) to 23% less than fully modeled (Michigan) (not shown). The effects of using total conditioned loading for individual tributaries may be greater than shown in Figure 5 depending on the magnitude of the residuals in the model (Figure 2). When using large spatial models, such as these SPARROW models, it is important to consider potential regional biases and local errors in interpreting the final model results.

SUMMARY AND CONCLUSIONS

Excessive nutrient loading has caused eutrophication problems in many areas of the Great Lakes. Previously developed SPARROW models had been used to describe where and from what sources the nutrients originate; however, those models only simulated export from the U.S. part of the Great Lakes Basin, were calibrated with potentially biased loads, and were calibrated using statistical techniques that may have resulted in not all calibration sites having equal influence on the final model. To address these issues, SPARROW models were developed for P and N for a 2002 base year for the binational Midcontinental Region of Canada and the U.S., including the entire area draining to the Great Lakes. Estimation of loading from throughout the entire Great Lakes system, including interlake and atmospheric inputs, enabled the origin of the entire external load and the sources of the load to each of the Great Lakes and all tributaries to be determined and compared. SPARROW models developed in this study were based on much smaller catchments (~2 km^2) than used in previous models enabling loads to be described over smaller areas, such as throughout watersheds of individual tributaries. The models were developed using a statistical approach to reduce the influence of nonuniformly distributed sites (thinning) and sites that are nested within basins of other downstream sites (weighting of residuals during calibration).

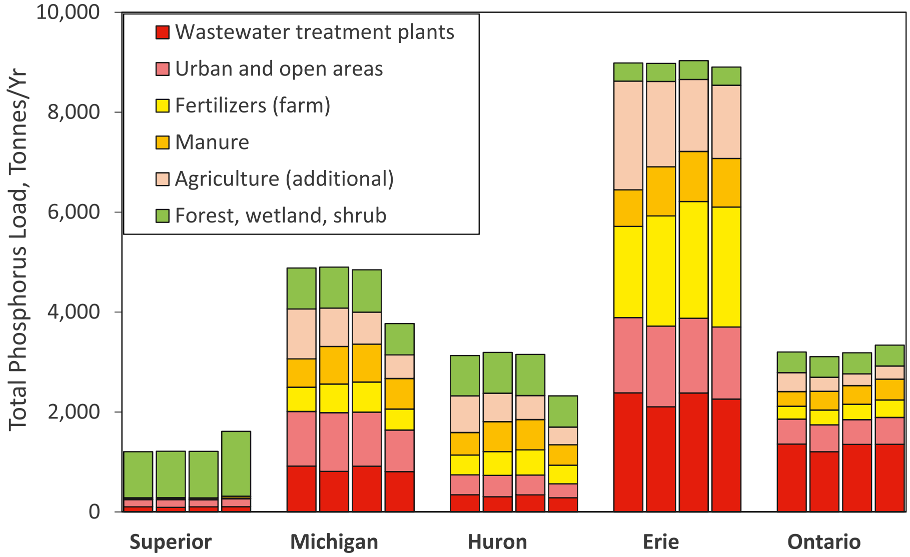

FIGURE 5. Total load and source allocation for each Great Lakes for preliminary and final SPARROW P models. For each lake, the first bar represents preliminary model results with all data considered, the second bar represents preliminary model results with data thinned to one site per 12-digit hydrologic unit code, the third bar represents full model results with data thinned and adjusted for nested basins, and the fourth bar represents loads adjusted to measured loads in the basin (loads are estimated by SPARROW only for unmonitored areas).

Phosphorus and Nitrogen loadings were determined for each of the Great Lakes. Overall loadings to each lake were relatively similar to that found in previous studies. Results of this study, however, provide a better spatial coverage of the loadings that can be used to prioritize where actions may have the largest effect and describe the importance of each P and N source that can be used to guide the type of actions needed to reduce loading. Inputs from WWTPs and urban areas were not as important of a P source as suggested in earlier SPARROW models but were still important sources of P (ranging from 11% for Superior to 37%–40% for Erie and Michigan).

While this paper focused on the Great Lakes Basin, results from Midcontinent SPARROW models are also useful for addressing water-quality issues throughout the entire modeled area. To allow scientists and resource managers easy access to all model results, an online mapping tool has been developed that allows users to map and download model results at a variety of scales anywhere in the Midcontinental Region, such as Lake Erie or the Rainy-Lake of the Woods watershed (https://sparrow.wim.usgs.gov/midcontinent-2002). Loads and yields can be spatially displayed for the entire modeled area or various subsets, all subareas can be ranked, and largest sources for each area can be displayed. All of the Midcontinent SPARROW model inputs and outputs are available online (Vouk et al. 2018b; Saad et al. 2019, respectively).

SPARROW models developed in this study represent the conditions and nutrient inputs similar to the early 2000s. Because of the time required by states and provinces to assemble monitoring results and the nutrient input information required for large spatial models, it is difficult to develop models that represent current-day practices. Ongoing studies that incorporate all of the approaches described in this paper are now being conducted with plans to update the models presented in this paper to nutrient inputs and practices similar to the early 2010s.

SUPPORTING INFORMATION

Additional supporting information may be found online under the Supporting Information tab for this article: Output from the midcontinent phosphorus and nitrogen SPARROW models are presented for: loads and yields to each of the Great Lakes (Table S1); loads and yields from each HUC8-sized basin (Table S2); and loads and yields for each tributary > 150 km^2 to each of the Great Lakes (Table S3). Tributaries to each Great Lake are ranked by relative loads and yields. All model results can be explored using the online mapping tool at this link: https://sparrow.wim.usgs.gov/midcontinent-2002. [Correction added on September 25, 2019, after first online publication: in the supporting information, the following sentence is added at the end of the paragraph, "All model results can be explored using the online mapping tool at this link: https://sparrow.wim.usgs.gov/midcontinent-2002".]

ACKNOWLEDGMENTS

Financial support for this study was provided by the International Joint Commission. The authors would like to thank Craig Johnston (USGS, deceased), Richard Burcher (National Research Council Canada), and John Gaiot (Ontario Ministry of Natural Resources) for assistance with developing spatial coverages for the SPARROW models. Any use of trade, firm, or product names is only for descriptive purposes and does not imply endorsement by the U.S. or Canadian Governments.

LITERATURE CITED

AAFC (Agriculture and Agri-Food Canada). 1996. "Soil Landscape of Canada Version 2.2." http://sis.agr.gc.ca/cansis/nsdb/slc/v2.2/index.html.

AAFC (Agriculture and Agri-Food Canada). 2001. "Interpolated Census of Agriculture to Soil Landscapes, Ecological Frameworks, and Drainage Areas of Canada." Documentation and User Guide. Version 1, 32 pp.

AAFC (Agriculture and Agri-Food Canada). 2002. "2001 Census of Agriculture." http://www.statcan.gc.ca/pub/95f0301x/.

AAFC (Agriculture and Agri-Food Canada). 2007. "2001 Census of Agriculture: Data Tables." http://www.statcan.gc.ca/pub/95f0301x/4151237-eng.htm.

Alexander, R.B.2015. "Advances in Quantifying Streamflow Variability Using Hierarchical Bayesian Methods with SPARROW." PhD diss., The Pennsylvania State University, College of Agricultural Sciences, p. 281.

Alexander, R.B., R.A. Smith, G.E. Schwarz, E.W. Boyer, J.V. Nolan, and J.W. Brakebill. 2008. "Differences in Phosphorus and Nitrogen Delivery to the Gulf of Mexico from the Mississippi River Basin." *Environmental Science and Technology* 42 (3): 822–30.

Amemiya, T. 1985. *Advanced Econometrics*. Cambridge, MA: Harvard University Press. ISBN 9780674005600, p. 536.

Beaulac, M.N., and K.H. Reckhow. 1982. "An Examination of Land Use — Nutrient Export Relationships." *Water Resources Bulletin* 18 (6): 1013–24.

Benoy, G.A., R.W. Jenkinson, D.M. Robertson, and D.A. Saad. 2016. "Nutrient Delivery to Lake Winnipeg from the Red — Assiniboine River Basin — A Binational Application of the SPARROW Model." *Canadian Water Resources Journal/Revue Canadienne des Ressources Hydriques* 41: 429–47. https://doi.org/10.1080/07011784.2016.1178601.

Centre for Topographic Information. 2009. "Land Cover, Circa 2000-Vector Data Product, Edition 1.0." Earth Sciences Sector, Natural Resources Canada, 21 pp.

Centre for Topographic Information. 2012. *CanVec: Data Product Specifications — Edition 1.1*. Sherbrooke, QC: Natural Resources Canada. ftp://ftp2.cits.nrcan.gc.ca/pub/canvec/doc/CanVec_product_specifications_v1_1.pdf.

Cohn, T.A. 2005. "Estimating Contaminant Loads in Rivers: An Application of Adjusted Maximum Likelihood to Type 1 Censored Data." *Water Resources Research* 41 (7): 1–13.

Dolan, D.M., and S.C. Chapra. 2012. "Great Lakes Total Phosphorus Revisited: 1. Loading Analysis and Update (1994–2008)." *Journal of Great Lakes Research* 38: 730–40.

Dove, A., and S.C. Chapra. 2015. "Long-Term Trends of Nutrients and Trophic resPonse Variables for the Great Lakes." *Limnology and Oceanography* 60: 696–721.

ECCC (Environment and Climate Change Canada). 2013. "The Great Lakes Nutrient Initiative." http://www.ec.gc.ca/default.asp?xml:lang=En&n=11A9E7E5-1&news=7D6BB6E6-F67C-4477-AFB7-19E6310A8E15.

ECCC (Environment and Climate Change Canada). 2016. "National Water Data Achive: HYDAT." https://www.canada.ca/en/environment-climate-change/services/water-overview/quantity/monitoring/survey/data-products-services/national-archive-hydat.html.

Encyclopedia Britannica. 2019. "Area and Volumes of the Great Lakes." https://www.britannica.com/topic/Areas-and-Volumes-of-the-Great-Lakes-1800353.

GLWQA (Great Lakes Water Quality Agreement). 1978. *Revised Great Lakes Water Quality Agreement of 1978, as Amended by Protocol November 18th, 1987*. Windsor, ON: International Joint Commission.

Hirsch, R.M. 2014. "Large Biases in Regression-Based Constituent Flux Estimates: Causes and Diagnostic Tools." *Journal of the American Water Resources Association* 50 (6): 1401–24.

Homer, C., J. Dewitz, J. Fry, M. Coan, N. Hossain, C. Larson, N. Herold, A. McKerrow, J.N. VanDriel, and J. Wickham. 2007. "Completion of the 2001 National Land Cover Database for the Conterminous United States." *Photogrammetric Engineering and Remote Sensing* 73 (4): 337–41.

Hong, B., D.P. Swaney, and R.W. Howarth. 2011. "A Toolbox for Calculating Net Anthropogenic Nitrogen Inputs (NANI)." *Environmental Modelling and Software* 26 (5): 623–33. https://doi.org/10.1016/j.envsoft.2010.11.012.

Hoos, A.B., S. Terziotti, G. McMahon, K. Savvas, K.C. Tighe, and R. Alkons-Wolinsky. 2008. Data to Support Statistical Modeling of Instream Nutrient Load Based on Watershed Attributes, Southeastern United States, 2002." *U.S. Geological Survey Open-File Report 2008–1163*, p. 50.

Jenkinson, R.W., and G.A. Benoy. 2015. *Red-Assiniboine Basin SPARROW Model Development Technical Document*. Ottawa, ON: National Research Council of Canada. http://ijc.org/files/publications/IJC%20RA%20Model%20Development%20Report%202015%20FINAL.pdf.

Lee, C.J., R.M. Hirsch, G.E. Schwarz, D.J. Holtschlag, S.D. Preston, C.G. Crawford, and A.V. Vecchia. 2016. "An Evaluation of Methods for Estimating Decadal Stream Loads." *Journal of Hydrology* 542: 185–203.

Maccoux, M.J., D.M. Dolan, and S.C. Chapra. 2013. "Chloride and Total Phosphorus Budgets for Green Bay, Lake Michigan." *Journal of Great Lakes Research* 39: 420–28.

Makarewicz, J.C., W.G. Booty, and G.S. Bowen. 2012. "Tributary Phosphorus Loading to Lake Ontario." *Journal of Great Lakes Research* 38: 14–20.

McCabe, G.J., and D.M. Wolock. 2011. "Independent Effects of Temperature and Precipitation on Modeled Runoff in the Conterminous United States." *Water Resources Research* 47 (11). https://doi.org/10.1029/2011WR010630.

McKenney, D.W., P. Papadopol, K.L. Campbell, K.M. Lawrence, and M.F. Hutchinson. 2006. "Spatial Models of Canada- and North America-Wide 1971/2000 Minimum and Maximum Temperature, Total Precipitation and Derived Bioclimatic Variables." Frontline Technical Notes 106, Natural Resources Canada, Great Lakes Forestry Centre, Sault St. Marie, ON, 9 pp.

McMahon, G., L. Tervelt, and W. Donehoo. 2007. "Methods for Estimating Annual Wastewater Nutrient Loads in the Southeastern United States." U.*S. Geological Survey Open-File Report 2007–1040*, p. 81.

Minns, C.K., M. Munawar, M.A. Koops, and E.S. Millard. 2011. "Long-Term Ecosystem Studies in the Bay of Quinte, Lake Ontario, 1972–2008: A Prospectus." *Aquatic Ecosystem Health and Management* 14: 3–8.

Moore, R.B., and T.G. Dewald. 2016. "The Road to NHDPlus — Advancements in Digital Stream Networks and Associated Catchments." *Journal of the American Water Resources Association* 52 (4): 890–900. https://doi.org/10.1111/1752-1688.12389.

Mueller, D.K., and J.M. Gronberg. 2013. "County-Level Estimates of Nitrogen and Phosphorus from Animal Manure for the

Conterminous United States, 2002." *U.S. Geological Survey Open-File Report Report 2013–1065*.

Nakagaki, N., M.E. Wieczorek, and S.L. Qi. 2016. "Estimates of Subsurface Tile Drainage Extent for the Conterminous United States, Early 1990s." *U.S. Geological Survey, data release*. https://doi.org/10.5066/F7RB72QS.

NASS (National Agriculture Statistics Service). 2004. *2002 Census of Agriculture — State and County Data, Volume 1: Geographic Area Series*. Washington, D.C.: U.S. Department of Agriculture. http://agcensus.mannlib.cornell.edu/AgCensus/censusParts.do?year=2002.

National High Impact Weather Laboratory. 2014. "The Canadian Precipitation Analysis (CaPA)." *Environment and Climate Change Canada*, p. 5. http://collaboration.cmc.ec.gc.ca/cmc/cmoi/product_guide/docs/lib/capa_information_leaflet_20141118_en.pdf.

Ontario Ministry of Agriculture Food and Rural Affairs. 2015. "Tile Drainage Area." Land Information Ontario, https://www.javacoeapp.lrc.gov.on.ca/geonetwork/srv/en/main.home.

Ontario Ministry of Natural Resources and Forestry. 2012. "Ontario Integrated Hydrology Data, 1.0 Initial Data Release: November 2012." Land Information Ontario. https://www.ontario.ca/data/ontario-integrated-hydrology-data.

PRISM (Parameter-elevation Regressions on Independent Slopes Model) Climate Group. 2009. "PRISM Climate Data." Oregon State University. http://prism.oregonstate.edu.

Qian, S.S., K.H. Reckhow, J. Zhai, and G. McMahon. 2005. "Nonlinear Regression Modeling of Nutrient Loads in Streams: A Bayesian Approach." *Water Resources Research* 41: W07012. https://doi.org/10.1029/2005WR003986.

Reckhow, K.H., M.N. Beaulac, and J.T. Simpson. 1980. "Modeling Phosphorus Loading and Lake Response under Uncertainty: A Manual and Compilation of Export Coefficients." *U.S. EPA Report No. EPA–440/5–80–011*, Office of Water Regulation, Criteria and Standards Division, U.S. Environmental Protection Agency, Washington, D.C.

Richards, R.P., I. Alameddine, J.D. Allan, D.B. Baker, N.S. Bosch, R. Confesor, J.V. DePinto, D.M. Dolan, J.M. Reutter, and D. Scavia. 2013. "Discussion: Nutrient Inputs to the Laurentian Great Lakes by Source and Watershed Estimated Using SPARROW Watershed Models by D.M. Robertson and D.A. Saad." *Journal of the American Water Resources Association* 49 (3): 715–24.

Robertson, D.M., and D.A. Saad. 2011. "Nutrient Inputs to the Laurentian Great Lakes by Source and Watershed Estimated Using SPARROW Watershed Models." *Journal of the American Water Resources Association* 47 (5): 1011–33. https://doi.org/10.1111/j.1752-1688.2011.00574.x.

Robertson, D.M., and D.A. Saad.2013. "SPARROW Models Used to Understand Nutrient Sources in the Mississippi/Atchafalaya River Basin." *Journal of Environmental Quality* 42 (5): 1422–40. https://doi.org/10.2134/jeq2013.02.0066.

Ruddy, B.C., D.L. Lorenz, and D.K. Mueller. 2006. "County-Level Estimates of Nutrient Inputs to the Land Surface of the Conterminous United States, 1982–2001." *U.S. Geological Survey Scientific Investigations Report 2006–5012*, p. 17.

Saad, D.A., G.A. Benoy, and D.M. Robertson. 2018. "Estimates of Long-Term Mean Annual Nutrient Loads Considered for Use in SPARROW Models of the Midcontinental Region of Canada and the United States, 2002 Base Year." *U.S. Geological Survey, Scientific Investigation Report 2018–5051*. http://pubs.er.usgs.gov/publication/sir20185051.

Saad, D.A., D.M. Robertson, G.A. Benoy, and I. Vouk. 2019. "SPARROW Model Simulated Nutrient Loads in Streams of the Midcontinental Region of Canada and the United States, 2002 Base Year." *U.S. Geological Survey, data release*. https://doi.org/10.5066/P9ODSQYB.

Saad, D.A., G.E. Schwarz, D.M. Robertson, and N.L. Booth. 2011. "A Multi-Agency Nutrient Dataset Used to Estimate Loads, Improve Monitoring Design, and Calibrate Regional Nutrient SPARROW Models." *Journal of the American Water Resources Association* 47 (5): 933–49.

Schwarz, G.E., A.B. Hoos, R.B. Alexander, and R.A. Smith. 2006. "The SPARROW Surface Water-Quality Model: Theory, Application and User Documentation." *U.S. Geological Survey Techniques and Methods*, Book 6, Chapter B3, Reston, VA.

Schwede, D.B., R.L. Dennis, and M.A. Bitz. 2009. "The Watershed Deposition Tool: A Tool for Incorporating Atmospheric Deposition in Water-Quality Analyses." *Journal of the American Water Resources Association* 45 (4): 973–85.

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

Smith, R.A., G.E. Schwarz, and R.B. Alexander. 1997. "Regional Interpretation of Water-Quality Monitoring Data." *Water Resources Research* 33: 2781–98.

Stenback, G.A., W.A. Crumpton, K.E. Schilling, and M.J. Helmers. 2011. "Rating Curve Estimation of Nutrient Loads in Iowa Rivers." *Journal of Hydrology* 396: 158–69.

Sugg, Z. 2007. *Assessing U.S. Farm Drainage: Can GIS Lead to Better Estimates of Subsurface Drainage Extent?* Washington D.C.: World Resources Institute. http://pdf.wri.org/assessing_farm_drainage.pdf.

U.S. Geological Survey. 2017. "USGS Water Data for the Nation." https://waterdata.usgs.gov/nwis.

USACE (United States Army Corp of Engineers). 2016. "National Inventory of Dams, 2016 NID Database: U.S. Army Corp of Engineers." https://catalog.data.gov/dataset/national-inventory-of-dams.

USEPA (U.S. Environmental Protection Agency). 2018a. "The Great Lakes: Facts and Figures." https://www.epa.gov/greatlakes/facts-and-figures.

USEPA (U.S. Environmental Protection Agency). 2018b. "CMAQ: The Community Multiscale Air Quality Modeling System." https://www.epa.gov/cmaq.

Vouk, I., R.S. Burcher, C.M. Johnston, R.W. Jenkinson, D.A. Saad, J.S. Gaiot, G.A. Benoy, and D.M. Robertson. 2018a. "Data and Procedures Used to Develop Inputs for Nutrient SPARROW Models throughout the Midcontinental Region of Canada and the United States." National Research Council of Canada. p. 39. plus appendices. https://doi.org/10.4224/23004810.

Vouk, I., R.S. Burcher, C.M. Johnston, R.W. Jenkinson, D.A. Saad, J.S. Gaiot, G.A. Benoy, D.M. Robertson, and M. Laitta. 2018b. *Binational Dataset Used to Develop Nutrient SPARROW Models for the Midcontinental Region of Canada and the United States*. National Research Council of Canada. https://doi.org/10.4224/300.0001.

Watson, S.B., C. Miller, G. Arhonditsis, G.L. Boyer, W. Carmichael, M.N. Charlton, R. Confesor, et al. 2016. "The Re-Eutrophication of Lake Erie: Harmful Algal Blooms and Hypoxia." *Harmful Algae* 56: 44–66.

Wellen, C., G.B. Arhonditsis, T. Labencki, and D. Boyd. 2014. "Application of the SPARROW Model in Watersheds with Limited Information: A Bayesian Assessment of the Model Uncertainty and the Value of Additional Monitoring." *Hydrological Processes* 28: 1260–83.

Wolock, D.M.1997. "STATSGO Soil Characteristics for the Conterminous United States." *U.S. Geological Survey Open-File Report-656*. https://pubs.er.usgs.gov/publication/ofr97656.

Wolock, D.M., and G.J. McCabe. 1999. "Estimates of Runoff Using Water-Balance and Atmospheric General Circulation Models." *Journal of the American Water Resources Association* 35 (6): 1341–50. https://doi.org/10.1111/j.1752-1688.1999.tb04219.x.

Wolock, D.M., and G.J. McCabe. 2018. "Water Balance Model Inputs and Outputs for the Conterminous United States, 1900–2015." U.S. Geological Survey Data Release. https://doi.org/10.5066/F71V5CWN.