Robertson, Dale M., and David A. Saad, 2013. Reply to Discussion — “Nutrient Inputs to the Laurentian Great Lakes by Source and Watershed Estimated Using SPARROW Watershed Models” by R. Peter Richards, Ibrahim Alameddine, J. David Allan, David B. Baker, Nathan S. Bosch, Remegio Confesor, Joseph V. DePinto, David M. Dolan, Jeffrey M. Reutter, and Donald Scavia. Journal of the American Water Resources Association (JAWRA) 49(3): 725-734. DOI: 10.1111/jawr.12060

Richards et al. (2013) recently expressed concerns about the accuracy and management relevance of the study by Robertson and Saad (2011), which employed the regional watershed modeling technique, SPAtially Referenced Regression On Watershed attributes model — SPARROW, to quantify the sources and transport of nutrients in streams of the Great Lakes, Mississippi, and Ohio River Basins (Upper Midwest, referred to as Major River Basin #3 — MRB3). Many of Richards et al.’s conclusions about the accuracy of the SPARROW results are based on analyses of eight stream monitoring sites near Lake Erie, where detailed measurements of nutrients have been collected on a daily or more frequent basis by the National Center for Water Quality Research (NCWQR) at Heidelberg University for more than 10 years. We agree with Richards et al.’s estimates of bias in the load predictions at these eight monitoring sites and we share their concerns about the importance of working to advance the methods of load estimation to reduce biases in watershed models. In the following sections, we respond to their specific criticisms and report our current understanding of the magnitude of the bias in the U.S. Geological Survey (USGS) load estimates and its likely effect on the assessment of nutrient loads and sources to the Great Lakes by Robertson and Saad (2011). We initially provide background information on the SPARROW model to provide context for discussion of the effects of that bias on model predictions and on the utility of the models for supporting management decisions. And, in a final section, we highlight on-going and future USGS activities directed at improving the reliability of the estimates of stream water-quality loads. Based on the information that we present here, we conclude that, despite evidence of biases in loads at the eight monitoring sites near Lake Erie, the nutrient assessment presented by Robertson and Saad (2011) provides a valid assessment of the nutrient sources and transport factors in streams and water-
sheds of the larger MRB3 region, including the United States (U.S.) portion of the Great Lakes Basin, and provides information that can assist with nutrient management in the region.

THE SPARROW WATER-QUALITY MODELING FRAMEWORK

SPARROW watershed models are designed to integrate large quantities of hydrologic and geospatial information to explain the spatial variation in water-quality conditions over large regions. Historically, water-quality data were collected at many sites throughout a region, but interpretation of those data was often complicated by the lack of a spatial framework to define the regional importance of each site. Furthermore, there was a poor linkage between the monitoring information and upstream sources of contaminants, thus complicating the identification of the relative importance of the sources. Many types of geospatial data that are related to water quality (e.g., land use) have become available over recent decades and that information has provided the opportunity to develop analytical methods to better utilize and interpret available water-quality data. SPARROW is one such method and is designed to comprehensively integrate a variety of types of geospatial data to provide a means of extrapolating and interpreting available monitoring information. The detailed spatial framework inherent in SPARROW models provides a means of predicting water quality across scales ranging from small drainages to large river basins that discharge to receiving waters. It is the first such method that integrates deterministic and statistical concepts to optimally estimate water-quality relations across large regions like the U.S. part of the Great Lakes Basin.

A critical aspect of SPARROW models is that they are based on mass balance and its predictions are constrained to optimally match measured water-quality loads as defined by monitoring data. Comprehensive geospatial data describing source contributions are combined with similar datasets describing geographic characteristics associated with transport processes (e.g., stream travel time). Mass contributions from sources are balanced with losses to optimally match water quality at monitoring locations such as those operated by the NCWQR. This optimal mass balance combined with spatially explicit estimates of source contributions forms the basis of predictions for all streams within a study region.

We believe that SPARROW models provide comprehensive tools for integrating many types of information and testing assumptions about the relative contributions of contaminants from watersheds and the relative importance of different types of contaminant sources. In fact, there are few modeling techniques that integrate more information to understand regional water quality. Furthermore, the mass-balance constraints inherent in SPARROW provide an opportunity to evaluate the role of watershed characteristics by statistically relating them to water-quality measurements collected at many sites located throughout the region. We believe that the comprehensiveness of information, and the integration of that information constrained to water-quality measurements lends credibility to the model results. Such results may challenge the previous understanding of water-quality conditions developed using simpler techniques. Those challenges should be given due consideration both for scientific evaluation and for water-quality management decisions. In fact, we would encourage that management decisions be based on an array of lines of evidence including monitoring data as well as model predictions. SPARROW results should clearly be part of that and can provide many types of useful information for management purposes.

LOAD ESTIMATES USED IN SPARROW MODELING

Given the mass-balance nature of SPARROW models, compiling information describing stream loads is a critical part of their development. Typically, to develop a SPARROW model, loading information is compiled from as many sites as possible so as to represent the full range of spatial variability across the study area. This activity often involves compiling data from many agencies that collect water-quality information in the region and evaluating those data for the potential to estimate a load with reasonable certainty. This approach offers multiple advantages including maximizing the utility of all the resources expended for monitoring, maximizing the statistical power of the models, and providing a broad interpretive context for the agencies collecting the data. However, this approach also adds some level of uncertainty in the SPARROW model. We chose to emphasize the benefits of compiling as much data as possible and addressing any added uncertainty through statistical evaluation as part of the calibration process.

Prior to selecting water-quality data from a particular site, all records were screened based on a set of criteria established to avoid sources of load-estimate error. The criteria were based on record length, number of samples, and proximity to the period of interest (see Robertson and Saad, 2011; Saad et al., 2011 and
the associated Supporting Information for details). Furthermore, even after the data were used for load estimation, additional screening criteria were applied to ensure that the load estimates were reliable. Sites for which the regression relation performed poorly (standard error >50%) were either reevaluated or not included in the model calibration. Overall this screening and load-estimate evaluation process resulted in total phosphorus (P) load estimates from 810 sites and total nitrogen (N) load estimates from 708 sites in the MRB3 region.

In any load-estimation effort, bias can come from a variety of causes, sometimes in spite of efforts to detect or mitigate it. Those causes can include mathematical retransformation from log space, violation of load-estimation model assumptions, and unrepresentative samples. USGS load-estimation techniques when properly implemented are designed to counter many of these sources of bias, but are not foolproof. Recent studies highlighting these facts have applied new diagnostics designed to identify load-estimate bias where it occurs (Stenback et al., 2011; Garrett, 2012; Moyer et al., 2012). The use of residuals plots and statistical diagnostics can identify potential bias problems in many cases. The USGS is currently using the NCWQQR records to further the understanding of load-estimation bias. In addition, two recent USGS reports (Garrett, 2012; Moyer et al., 2012) provide examples of load-estimation studies that focus on detecting and counteracting sources of load-estimate bias.

THE ACCURACY OF SPARROW LOAD ESTIMATES

We acknowledge the unique and important contribution that the detailed NCWQQR records reported by Richards et al. make to the understanding of chemical loadings in the streams near Lake Erie and more generally as a resource to support the development of more accurate load-estimation techniques. Because of the high frequency of samples, the NCWQQR records provide highly accurate measures of nutrient loads that can be reliably used to assess the accuracy of the USGS load estimates (Robertson and Saad, 2011) at these sites. We agree with Richards et al.’s estimates of the magnitude of the bias in the mean annual nutrient loads. Their estimates of bias in Fluxmaster-based loads for the eight sites were all less than 31% for P (median = 16%) and less than 36% for N (median = 9%), with the exception of one site with an over-estimation of 89% for N, and were generally larger for SPARROW load predictions at these sites, which are derived from the spatial model calibrated to the entire MRB3 region.

Richards et al. cite several examples among the eight monitoring sites for P and N where large biases in loads can result from poorly fitted regression models (tables 2 and 4 in Richards et al.). Regression-based load-estimation techniques, such as those used by Robertson and Saad (2011), are routinely used in water-resource assessments to estimate loads at stream monitoring sites, where the frequency of water-quality sampling is periodic (e.g., weekly, monthly, bimonthly) but flow measurements are relatively plentiful (e.g., 15 min, daily). Although the estimate of bias in the long-term mean load at the eight sites near Lake Erie were generally less than about 30%, Richards et al. report much larger biases in the daily loads. Biased estimates of annual and daily loads can arise when the assumptions of the estimated concentration-flow regression model (e.g., Fluxmaster) are strongly violated. This can include violations of the assumption of linearity for the relationship between log concentration and log flow and/or the assumption of homoscedastic residuals (i.e., constant variance in the residuals). Although we employed a rigorous set of criteria in our study to select monitoring sites with acceptable numbers of observations and levels of precision in the mean load (see Robertson and Saad, 2011 for details), we did not use methods that screened the sites explicitly for biases in the loads. New diagnostics have recently been developed to specifically detect load-estimate bias (Stenback et al., 2011) and these, along with standard regression diagnostics, should improve our capability for bias detection. However, a renewed focus on the issue of load-estimate uncertainty has revealed the extent of the problem and indicated the need for the broader water-research community to work toward further improving methods of detecting load-estimate bias and toward mitigating its causes. As a result, the USGS is beginning to take numerous steps to address these concerns as described in more detail in the final section of this response, including the use of new statistical methods to estimate loads (i.e., Weighted Regressions on Time, Discharge, and Season [WRTDS]; Hirsch et al., 2010; Moyer et al., 2012).

One approach that we are using to assess the potential for load bias for individual site records is the comparison of the sum of the measured loads on sampled days to the sum of the predicted loads on these days, expressed as a ratio of sums (see Saad et al., 2011; Garrett, 2012; Moyer et al., 2012). We evaluated the ratios of summed loads for the ~700-800 monitoring sites that were used to calibrate the MRB3 SPARROW model (as reported in table 1 of Saad et al., 2011). This evaluation helps to address the concern by Richards et al. that the estimates of...
bias for the eight NCWQR sites near Lake Erie may be indicative of biases in loads at monitoring sites throughout the MRB3 region. We find that the bias in the mean N load for the set of MRB3 monitoring sites is relatively small—typically less than 2% with most sites having a bias of less than 10% (ratio of measured to predicted load: median = 0.98; interquartile range [IQR] = 0.90-1.01; Table 1). We also find that the measured P loads were frequently underestimated at the MRB3 sites, although the bias in the mean load is typically only about 8%, with most sites having a bias of less than 25% (median = 1.08; IQR = 1.02-1.23). Note that our recent investigations using the more complete NCWQR records reveal that the bias ratio provides only an approximate measure of the true bias. Also, note that the ratios reported in Table 1 are based only on sampled days and do not reflect extrapolation-related biases that may result from model misspecification at high flows for which water-quality samples may be unavailable. By comparison, the accuracy (precision) of the loads computed with Fluxmaster at the MRB3 monitoring sites is estimated to be on average about ±25% and ±45% for N and P, respectively (Saad et al., 2011; estimates based on the 95% prediction intervals for the mean loads). Both random and systematic (bias) errors contribute to the estimates of load precision, although the effects of bias are likely to be only partially reflected in these estimates. The lower precision in the estimated P loads as compared with that for N loads can be explained by the higher variability in P concentrations in streams and complexities of the dissolved/particulate phases of phosphorus.

To assess the effect of load-estimate bias on model predictions, we evaluated the sensitivity of the SPARROW models based on our estimated bias-ratio measures (as reported in Table 1), and find that the predictions are only moderately sensitive to these levels of bias. For example, recalibrating the SPARROW models using bias-adjusted mean loads (Table 1) at the monitoring sites, we find that the P load for the U.S. portion of Lake Erie Basin increases by about 14% (increase from 4,610 to 5,270 MT/yr), with agricultural source shares increasing from 43 to 48% and point-source shares declining from 42 to 37% (urban nonpoint shares remain the same at about 10%). The total P load from the Maumee River Basin increases by about 17% (increase from 1,650 to 1,930 MT/yr), with agricultural source shares increasing from 63 to 67% and point-source shares declining from 26 to 22%. This increase in the relative importance of non-point-source contributions to point-source contributions was caused by increases in tributary loading and little change in most point-source contributions that were input directly to the lake. The results from a recalibrated N model were even less sensitive to the potential effects of the estimated bias in mean loads. Although the bias-adjusted mean loads only approximately capture the effects of the true bias on the SPARROW results, we conclude that the SPARROW predictions in watersheds of the MRB3 are not very sensitive to biases of the magnitude that we are able to quantify using the ratio metric. Therefore, this lack of sensitivity provides limited evidence that our prior model predictions give a sufficiently accurate description of the loads and sources throughout the MRB3 region. Nevertheless, we are continuing to evaluate and improve the methods for load estimation to address concerns about bias. These efforts include using the detailed NCWQR records to help guide our understanding of the performance of different methods and model specifications. The improvements that result from these investigations will be incorporated into future applications of the SPARROW model to the Great Lakes tributaries and to model applications in other regions of the U.S.

THE ACCURACY OF SPARROW SOURCE SHARES

Richards et al. also express concerns about the accuracy of the SPARROW estimates of source contri-
butions to Lake Erie. They argue that the share of P sources assigned to manure as compared to fertilizer was overstated and that nonpoint sources of P were underestimated and point sources were overestimated. As described in this section, we do not believe that these criticisms are supported by the information presented in their discussion. Their concerns are limited exclusively to P, and no criticisms were made of the SPARROW estimates of N source contributions.

Richards et al.’s criticism of SPARROW estimates of agricultural P sources was based on their prior analysis of nutrient sources in the Maumee River Basin. They argue that the relative proportions of fertilizer and manure P as estimated by SPARROW at the basin outlet (~60% fertilizer and ~40% manure, expressed as a percentage of the total agricultural stream load) should be similar to the proportions observed in applied fertilizers and manure at the field scale, where fertilizer inputs tend to dominate. We believe that their line of reasoning oversimplifies the chemical and physical processes that control the supply and delivery of P to downstream waters. Based on results from the MRB3 SPARROW model, we do find that fertilizer sources tend to dominate relative to manure sources (as suggested by Richards et al.), but at a somewhat lower level because P in manure is estimated to be delivered to streams at a higher rate than the P input from fertilizers. This higher delivery rate for manure is based on the higher estimated SPARROW delivery coefficient for manure from confined animals (applied to fields) as compared with that for farm fertilizer. This finding is consistent with recent studies (Stuntebeck et al., 2011), which show that much of the P lost from agricultural fields in the Midwest is associated with manure spreading, particularly on frozen ground during late winter and spring when loads are generally highest. Fertilizers are not generally applied during a period when runoff is most likely to occur. Environmental laws have only recently been established to stop the spreading of manure during the late winter and early spring (Midwest Environmental Advocates, 2012), so applications of manure during these seasons would have likely occurred during the modeled time period.

The separate tracking of the delivery of manure and fertilizer P to downstream waters is an important feature of the spatially explicit SPARROW model, which is designed to account for the location of different P mass inputs and the interactions of these inputs with delivery processes in soils, groundwater, streams, and reservoirs. This approach contrasts with more simplistic approaches that would deliver agricultural sources of P to downstream waters as a constant ratio of their application rate, or based on methods that simply infer that the larger input sources (e.g., fertilizer applications) are the more important sources in downstream waters (Richards et al., 2002). We contend that the SPARROW model predictions provide a more informed, physically based assessment of the importance of different types of agricultural sources on nutrient transport delivery to downstream waters.

Richards et al. further argue that SPARROW overestimates the contributions of point sources to Lake Erie loads because the results differ with those of prior studies that demonstrate that agricultural sources “dominate” nonpoint and total sources of loads to Lake Erie. They do not cite any specific numerical values to support their assertion that SPARROW overestimates point-source contributions, and they do not give any rationale as to why the prior studies should give more accurate predictions of sources than SPARROW. They instead cite as supporting “evidence” that prior Great Lakes studies have used “well-established approaches” for combining monitoring data with point-source inventories and other source information to predict loads to Lake Erie and its tributaries. They further claim that these assessments are “based primarily on land-use percentages (e.g., 76% agricultural, 12% urban),” with the conclusions supported by the Ohio Lake Erie Phosphorus Taskforce.

Although we agree with Richards et al. that there is a valuable body of scientific literature (consisting of both monitoring and modeling) in the Great Lakes region to which the authors of the commentary have been major contributors, this does not, without a thorough analysis and quantitative comparison, invalidate interpretations of sources using alternative models. In view of the uncertainties intrinsic to the estimation of source contributions in water samples (i.e., source contributions cannot be measured, but must be inferred from the use of watershed models), different modeling methods will invariably arrive at different conclusions about the nature of contaminant sources in watersheds because of underlying differences in data inputs, process assumptions, and various structural components of the models. We support the continued evaluation and comparison of alternative modeling approaches as this is essential to advance understanding of loads and source contributions. However, this will require more detailed comparisons and discussion of the methods than the information and analyses presented by Richards et al. SPARROW predictions as reported in Robertson and Saad (2011) indicate that point sources contribute ~40% of the total P loads to the U.S. part of Lake Erie, with agricultural loads being the predominant source in the tributary loads, and P loads from point sources are estimated to contribute
<25% of the upstream U.S. tributary loads to Lake Erie. According to SPARROW results, the relative contributions of point sources in Great Lakes loadings vary spatially but are typically smaller contributors than agricultural sources in the upstream watershed loads (e.g., Maumee River), whereas they are a larger contributor in the direct inputs from near-shoreline sources where many point sources are located. These estimates are based on SPARROW’s mass-balance accounting of point and nonpoint sources, which accounts for source interactions with watershed processes that influence transport and delivery. As we indicated earlier, our approach differs markedly with the approaches advocated by Richards et al. that employ land use and fertilizer and manure proportions (without accounting for watershed processes) as direct indicators of the expected nutrient-source contributions in downstream waters entering Lake Erie.

One additional source-related topic that was raised by Richards et al. is their concern that the SPARROW point-source loads are overestimated by a factor of 1.07 (a coefficient that is multiplied by the point-source loads in the calibrated SPARROW P model of Robertson and Saad, 2011). The inclusion of this factor is explained by our use of a mass-balance model to infer the contributions of P from point sources throughout the MRB3 region, based not only on the best available estimates of the P loadings in point-source effluent but also the outcome of comparing these loadings with monitored stream loads and the loads contributed by other watershed sources after accounting for interactions with transport processes. The adjustment factor of 7% results from a mass-balance assessment at the watershed scale that is inclusive of all nutrient sources. We maintain that this approach provides a more informed and accurate estimate of point-source contributions in MRB3 streams than a method that assumes that the individual reported loads from point-source facilities are measured without error. The advantage of our approach is that it recognizes that mean annual estimates of point-source loadings from individual facilities have large uncertainties that reflect various errors in measured concentrations, flows, and assumptions about the location and routing of effluent in facility pipes (Maupin and Ivahnenko, 2011). Moreover, it is informative that the factor of 1.07 (with standard error = 0.14) is actually statistically indistinguishable from a value of 1.0, indicating the magnitude of statistical uncertainty in our ability to estimate point-source loads. This serves as a further reminder that there are uncertainties in the water-resources community’s ability to use state-of-the-art methods to estimate nutrient sources, which contrasts with the assertion that point-source data should be assumed to provide estimates of loads with near perfect certainty.

ADDITIONAL CHARACTERISTICS OF SPARROW MODELS

Richards et al. raise several additional concerns about the limited scope of our study, which are readily explained by the objectives that were originally identified for our analysis. It is important to recognize that this study (Robertson and Saad, 2011) was part of a large nationwide investigation of the sources of nutrients in U.S. river basins (Preston et al., 2011) — one of the most comprehensive to date, which restricted the spatial and temporal scope of our modeling analysis of Great Lake tributaries in several ways.

First, our study only modeled the P and N loads to the Great Lakes from U.S. sources and land areas. Our study benefited from a comprehensive compilation of U.S. national data on land use, climate, soils, river networks, point-source effluent loads, river monitoring measurements, and other geospatial properties (see Preston et al., 2011). To have compiled these types of geospatial data for the Canadian drainage and integrated the data into our modeling framework would have required far more resources and time than were available for the study. In fact, we are currently collaborating with Canadian scientists now to complete joint U.S.-Canadian SPARROW models of P and N, which have revealed the challenges that exist in merging these data into a single model infrastructure; examples include differences in land-use classifications and in river and drainage network resolution. We agree with Richards et al. that a comprehensive Great Lakes model is a worthy objective, but the initial version of this — the U.S. Great Lakes model presented by Robertson and Saad (2011) — is an important and informative first step in that direction. The current model was not intended to be suitable for managing inputs to the entire Great Lakes, but is only relevant to advancing understanding of P and N sources and delivery from U.S. watersheds. We regret any misinterpretation of that objective, although it is clearly stated in Robertson and Saad (2011).

Second, the SPARROW loads are reported for a base year of 2002. Richards et al. is critical of our use of 2002 as a base year and argue for a more contemporaneous date for the model calibration so that the model predictions are relevant for current management. We have two responses. (1) There is often a lag between the published date of a modeling study and the time domain described by the model because con-
temporaneous data are difficult to obtain in a timely manner, owing to delays in laboratory analyses of water-quality samples, the periodic publishing schedules of certain state and federal data, and the complexities of assembling large geospatial datasets. For example, the Census of Agriculture by the U.S. Department of Agriculture is only published every five years (e.g., 1997, 2002, 2007) and the USGS National Land Cover Database (NLCD) is published every 10 years. Given the complexity of the national scope of the USGS SPARROW study of nutrients, in which Robertson and Saad (2011) participated, the selected base year of 2002 was the most current year that could accommodate the retrieval and analyses required of the datasets. (2) The time-averaged version of the SPARROW model is designed as a steady-state model of long-term mean annual nutrient conditions. Provided that dramatic changes have not occurred in nutrient sources and average climatic conditions since 2002, the model predictions of long-term average loads and source contributions should continue to be relevant beyond the stated base year as a description of the average response of stream conditions to watershed sources and processes and can be used to assist with current management decisions. Nevertheless, the USGS advocates the use of SPARROW predictions as an initial set of information to inform water-resource management that should be supplemented with more detailed and dynamic use of local stream monitoring data and modeling results, so that more current water-quality conditions can be evaluated.

Third, the SPARROW loads are normalized for long-term variations in streamflow. Richards et al. expressed concerns about our reporting of flow-normalized nutrient loads that are reflective of “typical” hydrology rather than reporting annual loads that show annual variability in flow conditions. They argue that “from a management perspective, ... the extreme loads are what is most important, especially if they occur close together in time.” In fact, the extreme loads and measures of the annual variations in load provide only a partial answer to the information needs of water managers. These measures are important to understanding the downstream effect of the actual magnitude of the annual variations in loads, which are caused by both natural and anthropogenic factors. This information is highly relevant to understanding, for example, the effects of loads on the ecology of receiving water bodies. By contrast, the flow-normalized, long-term, average-annual load, as reported by SPARROW, provides relevant information to address the question of what upstream anthropogenic and natural sources (controlling for the effects of interannual variations in flow) are responsible for the delivery of nutrients to receiving water bodies. This feature of SPARROW, together with the designation of a common base year, ensures that the nutrient loads in the mass-balance calculations are representative of long-term hydrologic variability during a consistent time period across monitoring sites so that the model provides robust estimates of the nutrient sources to streams and the major processes that govern the mean rates of nutrient removal and transport in watersheds. This approach offers a highly relevant perspective that can assist in the management of the sources of nutrients and other contaminants in watersheds. Indeed, the flow-normalized version of the SPARROW model has been used widely in the U.S. as a tool to assist water-resource managers with quantifying anthropogenic and natural sources, including in the Chesapeake Bay (Ator et al., 2011), Mississippi River Basin (Alexander et al., 2008), Colorado River Basin (Anning et al., 2007), and Southeastern U.S. (Hoos and McMahon, 2009).

Finally, SPARROW results were used to describe total loading and source allocations for the U.S. part of each Great Lake basin. Although SPARROW models are typically developed for large geographic areas, output from the models can be used to describe mean annual loads/yields and estimate the relative importance of nutrient sources over various geographic scales. In the Great Lakes Basin, there are ~160 tributaries with drainage areas >150 km², and ~100 eight-digit Hydrologic Unit Code basins (HUC8 basins; Seaber et al., 1987), making it impossible to describe loads, yields, and sources in all of them in a concise way. For that reason, Robertson and Saad (2011) summarized model results for each of the five Great Lakes, drawing conclusions based on that scale. Information for each tributary and HUC8 were also made available in the Supplementary Material included with Robertson and Saad (2011), and are available with various online tools, including the SPARROW Decision Support System (Booth et al., 2011; http://cida.usgs.gov/sparrow/), and SPARROW Mapper (http://wim.usgs.gov/Sparrow/SparrowMapper.html#). These tools enable scientists and managers to quantify and compare contributions from parts of the basins of each Great Lake and simulate the effects on stream loads of changes in nutrient-source contributions.

**FUTURE PLANS IN SPARROW MODELING IN THE GREAT LAKES BASIN**

The SPARROW model, like most other models, continues to evolve. Richards et al. suggest several enhancements in future SPARROW modeling efforts.
that would improve load estimates for Lake Erie. They include: (1) incorporate the Canadian portion of the Great Lakes watershed into the study area, (2) improve the accuracy of monitored load estimates, (3) use a more up-to-date base year, and (4) include different nutrient species (such as dissolved or bioavailable P) in the models. Suggestions (1), (2), and (3), as well as other improvements, are being incorporated into new models already under construction as described in more detail below. Research is also underway to develop SPARROW models that incorporate multiple nutrient species.

The spatial extent of the MRB3 SPARROW models was restricted to the U.S. portion of the Great Lakes, Upper Mississippi, Ohio, and Red River Basins (MRB3) because consistent geospatial data were not available from the Canadian portion of the Basin. Recently, however, collaborative work with Canadian scientists has begun to develop consistent international geospatial datasets that are required to develop SPARROW models for the entire Great Lakes Basin. Completion of those datasets will enable the development of SPARROW models for the entire Basin and the evaluation of loading and source contributions from the entire watershed of each lake.

Bias metrics have been added to the Fluxmaster program (Schwarz et al., 2006) used in SPARROW modeling and are being added to the USGS LOADEST program (http://water.usgs.gov/software/loa/dest/), which has been used in other USGS load studies (such as Garrett, 2012). As an additional diagnostic, we continue to support the careful inspection of standard residual plots as a means to identify model misspecification and nonconstant variance problems, which are believed to be the primary causes of biased load estimates. For example, these diagnostics have been successfully used together with additional explanatory variables (e.g., hysteresis, antecedent flow) to obtain improved LOADEST fits to water-quality data with less biased estimates of load at monitoring sites in Iowa rivers (Garrett, 2012). A newly developed load-estimation method, WRTDS (Hirsch et al., 2010) deals with problems of model misspecification and nonconstant variance through the use of statistical-smoothing techniques and has been shown to resolve load-estimation bias problems in many cases. Concerns about load-estimation bias recently led to the reevaluation and redesign of the methods used to estimate loads in the major tributaries to the Chesapeake Bay (Moyer et al., 2012) using the WRTDS method. As these activities illustrate, the USGS recognizes the issue of bias in load estimation as a highly important topic that has broad implications for the accurate use of historical water-quality records for research and water-resource management, including Total Maximum Daily Load assessments.

As load-estimation methods undergo continued improvements, users of the technology will continue to face challenges in identifying the appropriate criteria to select monitoring sites with accurate load estimates for calibrating watershed models. Monitoring records with the sampling frequencies of the NCWQR sites (at least one measurement per day over >10 years) for which the “true” load is known, are exceedingly rare. Less frequently sampled records, such as those evaluated in the Robertson and Saad (2011) study will continue to be essential to provide estimates of loads that can inform research and water management. For example, a recent evaluation of water-quality data collected by 72 federal, state, and local agencies at >100,000 monitoring sites in the conterminous U.S. over the period 1970-2006 (Saad et al., 2011), based on data retrievals from the USGS National Water Information System, the U.S. Environmental Protection Agency Storage and Retrieval data warehouse, and other data systems, such as that of NCWQR, found that only 13 sites had nutrient measurements collected with a frequency of more than once every five days over the period of record. Fewer than ~50 sites had >500 samples collected during the period of record and only 14 sites had a total sample size >1,000 during the period of record. By contrast, as many as 2,700 sites were identified (Saad et al., 2011) as having a minimal quantity of data to support load-model calibration (i.e., >25 samples over two years that were paired with daily flow data). Therefore, obtaining the load data needed to calibrate SPARROW or other watershed models often requires a tradeoff in site selection with an eye to balancing the accuracy of loads over space and time. The endpoints of the tradeoff can be expressed as either (1) opting to use only highly reliable load data in model calibrations based on selecting a few, very frequently sampled monitoring sites, which may have the disadvantage of restricting the description of spatial variation across the modeled domain, or (2) opting to use less reliable load data in model calibrations based on selecting many, less frequently monitored sites, which has the advantage of broadening the spatial description of load variations across the modeled domain. In the regional scale models employed by Robertson and Saad (2011), as in many applications of SPARROW to large regional basins, the latter option was generally given preference over the former, such that the model was developed to provide a more comprehensive spatial representation of sources and processes throughout the entire MRB3 region, of which the Great Lakes was a subregional watershed. We anticipate that the criteria needed to select site records that are likely to provide load estimates with low bias will become increasingly refined in future SPARROW studies to incorporate improved under-
standing and specific measures of load bias similar to those already being applied in USGS load studies (Garrett, 2012; Moyer et al., 2012).

The SPARROW models used to describe loading to the Great Lakes were based on nutrient inputs and land use similar to 2002 because this was the year with the most recently available water-quality data and explanatory geospatial data for the entire study area. Future SPARROW models will be developed based on more current information as it becomes available. However, prior to developing these more up-to-date models, it is already possible to simulate the effects of changes in specific inputs that can be quantified using the SPARROW Decision Support System (Booth et al., 2011; http://cida.usgs.gov/sparrow/). For example, the Metropolitan Council Environmental Services Metro wastewater treatment plant near St. Paul, Minnesota, has dramatically decreased P in their effluent by ~90% since 2002 (Heiskary and Wasley, 2010). The effects of these changes can currently be evaluated using the SPARROW Decision Support System, as can any implemented or expected change in nutrient-source contributions. In that way, management changes can be simulated and assessed as more current data and modeling results are being compiled.

The MRB3 SPARROW models were constructed with water-flow paths defined by streams and reservoirs included in the enhanced RF1 network. SPARROW models are now being developed with water-flow paths and reach catchments delineated with the National Hydrography Dataset Plus stream-reach coverage, which will enable improved spatial descriptions of where and from what sources nutrients originate.

Results from SPARROW models developed for the U.S. portion of the Upper Midwest (MRB3) describe the spatial variability in P and N loads and sources for the U.S. part of the Great Lakes Basin (Robertson and Saad, 2011). Results from these and future SPARROW models can assist managers in prioritizing areas for nutrient-management efforts by providing a basis for ranking areas according to their relative contributions. They can also assist managers in making decisions regarding the types of actions that would be most beneficial by providing a basis for identifying the most important sources in specific areas. By implementing the most appropriate actions in the most influential areas, it may be possible to reduce nutrient loading to the Great Lakes and thus mitigate eutrophication-related problems.

ACKNOWLEDGMENTS

We would like to thank Rich Alexander, Steve Preston, Bob Hirsch, Bob Gilliom, and Mike Woodside, all from the U.S. Geological Survey, for their input and editorial suggestions to this manuscript.

LITERATURE CITED

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):822-830.

Anning, D.W., N.J. Bauch, S.J. Gerner, M.E. Flynn, S.N. Hamlin, S.J. Moore, D.H. Schaefer, S.K. Anderholm, and L.E. Spangler, 2007. Revised 2010. Dissolved Solids in Basin-Fill Aquifers and Streams in the Southwestern United States. U.S. Geological Survey Scientific Investigations Report 2006-5315, 168 pp.

Ator, S.W., J.W. Brakebill, and J.D. Blomquist, 2011. Sources, Fate, and Transport of Nitrogen and Phosphorus in the Chesapeake Bay Watershed: An Empirical Model. U.S. Geological Survey Scientific Investigations Report 2011-5167, 27 pp.

Booth, N.L., E.J. Everman, I.L. Kuo, L. Sprague, and L. Murphy, 2011. A Web-Based Decision Support System for Assessing Regional Water-Quality Conditions and Management Actions. Journal of the American Water Resources Association 47(5): 1136-1150, doi: 10.1111/j.1752-1688.2011.00573.x.

Garrett, J.D., 2012, Concentrations, Loads, and Yields of Selected Constituents from Major Tributaries of the Mississippi and Missouri Rivers in Iowa, Water Years 2004-2008. U.S. Geological Survey, Scientific Investigations Report 2012-5240, 61 pp. http://pubs.usgs.gov/sir/2012/5240/.

Heiskary, S. and D. Wasley, 2010. Mississippi River Pools 1 Through 8: Developing River, Pool and Lake Popen Eutrophication Criteria. MPCA, St. Paul, Minnesota, 81 pp.

Hirsch, R.M., D.L. Moyer, and S.A. Archfield, 2010. Weighted Regressions on Time, Discharge, and Season (WRTDS), with an Application to Chesapeake Bay River Inputs. Journal of the American Water Resources Association 46(5):857-880.

Hoos, A.B. and G. McMahon, 2009. Spatial Analysis of Instream Nitrogen Loads and Factors Controlling Nitrogen Delivery to Streams in the Southeastern United States Using Spatially Referenced Regression on Watershed Attributes (SPARROW) and Regional Classification Frameworks. Hydrological Processes 23:2275-2294, doi: 10.1002/hyp.7323.

Maupin, M.A. and T. Ivahnenko, 2011. Nutrient Loadings to Streams of the Continental United States from Municipal and Industrial Effluent. Journal of the American Water Resources Association 47(5):950-964, doi: 10.1111/j.1752-1688.2011.00576.x.

Midwest Environmental Advocates, 2012. Manure Spreading on Frozen Ground Banned. http://midwestadvocates.org/issues-actions/actions/manure-spreading-on-frozen-ground-banned/, accessed June 2012.

Moyer, D.L., R.M. Hirsch, and K.E. Hyer, 2012. Comparison of Two Regression-Based Approaches for Determining Nutrient and Sediment Fluxes and Trends in the Chesapeake Bay Watershed. U.S. Geological Survey Scientific Investigations Report 2012-5244, 118 pp. http://pubs.usgs.gov/sir/2012/5244.

Preston, S.D., R.B. Alexander, G.E. Schwarz, and C.G. Crawford, 2011. Factors Affecting Stream Nutrient Loads: A Synthesis of Regional SPARROW Model Results for the Continental United States. Journal of the American Water Resources Association 47(5):891-915, doi: 10.1111/j.1752-1688.2011.00577.x.

Richards, R.P., I. Alameddine, J.D. Allun, 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):715-724.
Richards, R.P., D.B. Baker, and D.J. Eckert, 2002. Trends in Agriculture in the LEASEQ Watersheds, 1975–1995. Journal of Environmental Quality 31:17–24.

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):1011-1033, doi: 10.1111/j.1752-1688.2011.00574.x.

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):933-949, doi: 10.1111/j.1752-1688.2011.00575.x.

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, Section B, Chapter 3.

Seaber, P.R., F.P. Kapinos, and G.L. Knapp, 1987. Hydrologic Unit Maps. U.S. Geological Survey Water-Supply Paper 2294, 63 pp. http://water.usgs.gov/GIS/huc.html, accessed December 2012.

Stenback, G.A., W.G. Crumpton, K.E. Schilling, and M.J. Helmers, 2011. Rating Curve Estimation of Nutrient Loads in Iowa Streams. Journal of Hydrology 396:158-169.

Stuntebeck, T.D., M.J. Komiskey, M.C. Peppler, D.W. Owens, and D.R. Frame, 2011. Precipitation-Runoff Relations and Water-Quality Characteristics at Edge-of-Field Stations, Discovery Farms and Pioneer Farm, Wisconsin 2003–8. U.S. Geological Survey Scientific Investigations Report 2011–5008, 46 pp.