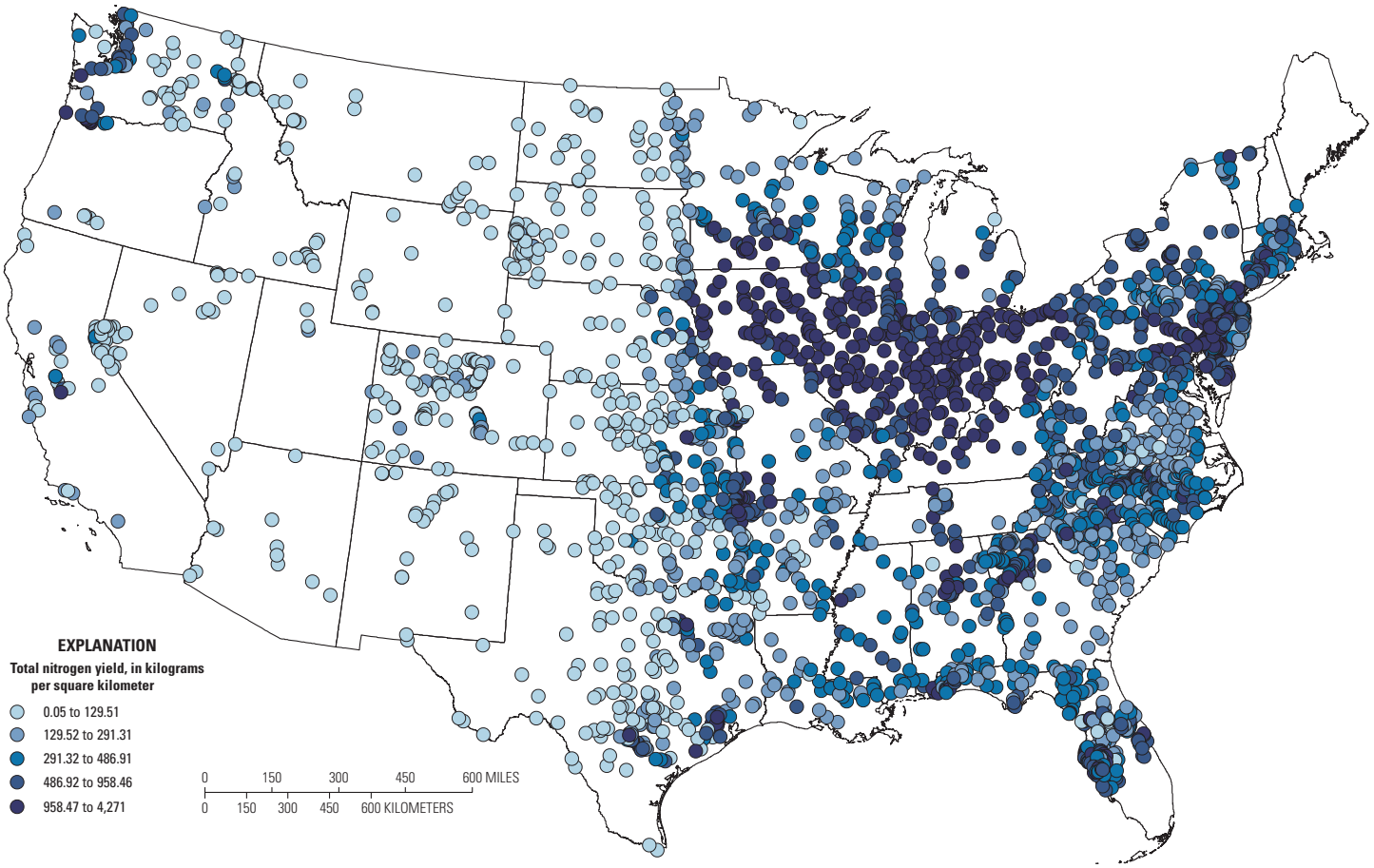
USGS
science for a changing world

Estimates of Long-Term Mean Daily Streamflow and Annual Nutrient and Suspended-Sediment Loads Considered for Use in Regional SPARROW Models of the Conterminous United States, 2012 Base Year



Scientific Investigation Report 2019-5069

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

Cover. Total nitrogen yield for stream sites considered for inclusion in 2012 regional Spatially Referenced Regressions on Watershed Attributes (SPARROW) nutrient models.

Estimates of Long-Term Mean Daily Streamflow and Annual Nutrient and Suspended-Sediment Loads Considered for Use in Regional SPARROW Models of the Conterminous United States, 2012 Base Year

By David A. Saad, Gregory E. Schwarz, Denise M. Argue, David W. Anning, Scott W. Ator, Anne B. Hoos, Stephen D. Preston, Dale M. Robertson, and Daniel R. Wise

Scientific Investigations Report 2019–5069

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

U.S. Department of the Interior
DAVID BERNHARDT, Secretary

U.S. Geological Survey
James F. Reilly II, Director

U.S. Geological Survey, Reston, Virginia: 2019

Suggested citation:
Saad, D.A., Schwarz, G.E., Argue, D.M., Anning, D.W., Ator, S.W., Hoos, A.B., Preston, S.D., Robertson, D.M., and Wise, D.R., 2019, Estimates of long-term mean daily streamflow and annual nutrient and suspended-sediment loads considered for use in regional SPARROW models of the conterminous United States, 2012 base year: U.S. Geological Survey Scientific Investigations Report 2019–5069, 51 p., https://doi.org/10.3133/sir20195069.

ISSN 2328-0328 (online)

Foreword

Sustaining the quality of the Nation's water resources and the health of our diverse ecosystems depends on the availability of sound water-resources data and information to develop effective, science-based policies. Effective management of water resources also brings more certainty and efficiency to important economic sectors. Taken together, these actions lead to immediate and long-term economic, social, and environmental benefits that make a difference to the lives of the almost 400 million people projected to live in the United States by 2050.

In 1991, Congress established the National Water-Quality Assessment (NAWQA) to address where, when, why, and how the Nation's water quality has changed, or is likely to change in the future, in response to human activities and natural factors. Since then, NAWQA has been a leading source of scientific data and knowledge used by national, regional, State, and local agencies to develop science-based policies and management strategies to improve and protect water resources used for drinking water, recreation, irrigation, energy development, and ecosystem needs (https://water.usgs.gov/nawqa/applications/). Plans for the third decade of NAWQA (2013–21) address priority water-quality issues and science needs identified by NAWQA stakeholders, such as the Advisory Committee on Water Information and the National Research Council. The plans are designed to meet increasing challenges related to population growth, increasing needs for clean water, and changing land-use and weather patterns.

Federal, State, and local agencies have invested billions of dollars to reduce the amount of pollution entering rivers and streams that millions of Americans rely on for a variety of water needs and biota rely on for habitat. Understanding the sources and transport of pollution is crucial for designing strategies to improve water quality. The U.S. Geological Survey's (USGS) Spatially Referenced Regressions on Watershed Attributes (SPARROW) model was developed to aid in the understanding of sources and transport of pollution across large spatial scales. The SPARROW model is estimated by statistically relating watershed sources and transport-related properties to monitoring-based water-quality load estimates. This report describes the water-quality and streamflow monitoring data, methods of estimating streamflow and loads, and estimates considered for use in regional SPARROW models (using 2012 as the base year). Monitoring data used in streamflow and load estimates were collected by more than 130 organizations from approximately 5,200 streamflow and 4,000 water-quality sites and cover the period from 1999 through 2014.

The authors hope this publication will provide insights and information to meet water-resource needs and will foster increased citizen awareness and involvement in the protection and restoration of our Nation's waters. The information in this report is intended primarily for those interested or involved in resource management and protection, conservation, regulation, and policy making at the regional and national levels.

Donald W. Cline
Associate Director for Water
U.S. Geological Survey

Contents

Figures

Tables

Conversion Factors

International System of Units to U.S. customary units

Multiply	By	To obtain
meter (m)	3.281	foot (ft)
kilometer (km)	0.6214	mile (mi)
square kilometer (km^2)	0.3861	square mile (mi^2)
liter (L)	0.2642	gallon (gal)
cubic meter per second (m^3/s)	35.31	cubic foot per second (ft^3/s)
millimeter per year (mm/yr)	0.03937	inch per year (in/yr)
kilogram per square kilometer (kg/km^2)	5.70995	pound per square mile (lb/mi^2)
metric ton per square kilometer (t/km^2)	0.000175	pound per square mile (lb/mi^2)
kilogram (kg)	2.205	pound avoirdupois (lb)

Abbreviations

BRE	Beale ratio estimator
F5K	Fluxmaster-K five-parameter regression
NAWQA	National Water-Quality Assessment
NHDPlus	National Hydrography Dataset Plus
NWIS	National Water Information System
SPARROW	Spatially Referenced Regressions on Watershed Attributes
SS	suspended sediment
STORET	Storage and Retrieval System
TN	total nitrogen
TP	total phosphorus
TSS	total suspended solids
USGS	U.S. Geological Survey
WRTDS	Weighted Regressions on Time, Discharge, and Season

Estimates of Long-Term Mean Daily Streamflow and Annual Nutrient and Suspended-Sediment Loads Considered for Use in Regional SPARROW Models of the Conterminous United States, 2012 Base Year

By David A. Saad, Gregory E. Schwarz, Denise M. Argue, David W. Anning, Scott W. Ator, Anne B. Hoos, Stephen D. Preston, Dale M. Robertson, and Daniel R. Wise

Abstract

Streamflow, nutrient, and sediment concentration data needed to estimate long-term mean daily streamflow and annual constituent loads were compiled from Federal, State, Tribal, and regional agencies, universities, and nongovernmental organizations. The streamflow and loads are used to develop Spatially Referenced Regressions on Watershed Attributes (SPARROW) models. SPARROW models help describe the distribution, sources, and transport of streamflow, nutrients, and sediment in streams throughout five regions of the conterminous United States. After the data were screened, approximately 5,200 streamflow, 3,000 sediment, and 3,300 nutrient sites, sampled by 137 agencies and organizations were identified as having suitable data for calculating the long-term mean daily streamflow and annual nutrient and sediment loads required for SPARROW model estimation. These sites are representative of a wide range in terms of watershed size, contaminant source types, and land-use and other important watershed characteristics. The methods used to estimate long-term mean annual loads include the Beale ratio estimator and Fluxmaster regression method with Kalman smoothing.

Introduction

Excessive nutrients and sediment are a persistent problem in streams and estuaries throughout the United States (Howarth and others, 1996; Carpenter and others, 1998; U.S. Environmental Protection Agency, 2000a; Howarth and others, 2002; Phillips, 2002; Bricker and others, 2007; Heimann, Sprague, and Blevins, 2011; Lee and Glysson, 2013). Excessive loading of phosphorus and nitrogen from anthropogenic sources has been linked to excessive macrophyte and phytoplankton growth in rivers and streams (U.S. Environmental Protection Agency, 1998), harmful algal blooms in parts of the Great Lakes (Michalak and others, 2013), and hypoxia in the Gulf of Mexico (U.S. Environmental Protection Agency, 2007). The U.S. Environmental Protection Agency identified sediment as the most common impairment to rivers and streams in the United States (U.S. Environmental Protection Agency, 2000b). To help improve understanding of the distribution, origins, and sources of these constituents, nutrients (total nitrogen [TN] and total phosphorus [TP]), and suspended sediment (SS), Spatially Referenced Regressions on Watershed Attributes (SPARROW) water-quality models are planned to be developed for five regions (Midwest, Northeast, Pacific, Southeast, and Southwest) of the conterminous United States (fig. 1), in part by using the load estimates described in this report.

SPARROW models simulate long-term mean annual constituent loads in streams and rivers over large geographic areas (Smith and others, 1997; Schwarz and others, 2006; Preston and others, 2009). The models describe the distribution of loads and yields and can be used to help understand the origin and transport of water-quality constituents. The regional SPARROW models will use readily available source and watershed characteristic spatial datasets, a digital representation of the stream network based on the National Hydrography Dataset Plus (NHDPlus) version 2 (Moore and Dewald, 2016), and long-term (1999-2014) mean daily streamflow and annual loads described in this report. Some functions of SPARROW models rely on estimates of long-term mean daily streamflow for each stream reach. SPARROW models designed to simulate long-term mean daily streamflows are also planned to generate a set of streamflow estimates, for each model region, that are coincident with the long-term mean constituent load estimates.

SPARROW models are estimated by relating long-term mean estimates of flow and load to the upstream watershed characteristics of the flow and load monitoring locations. Estimates of long-term mean flows and constituent loads require extended periods of data collection. A large set of flow and load monitoring sites, representing a wide range of watershed characteristics, is required to develop the model objectives (commonly referred to as “calibration targets”) used in the regional SPARROW models. Data from numerous water

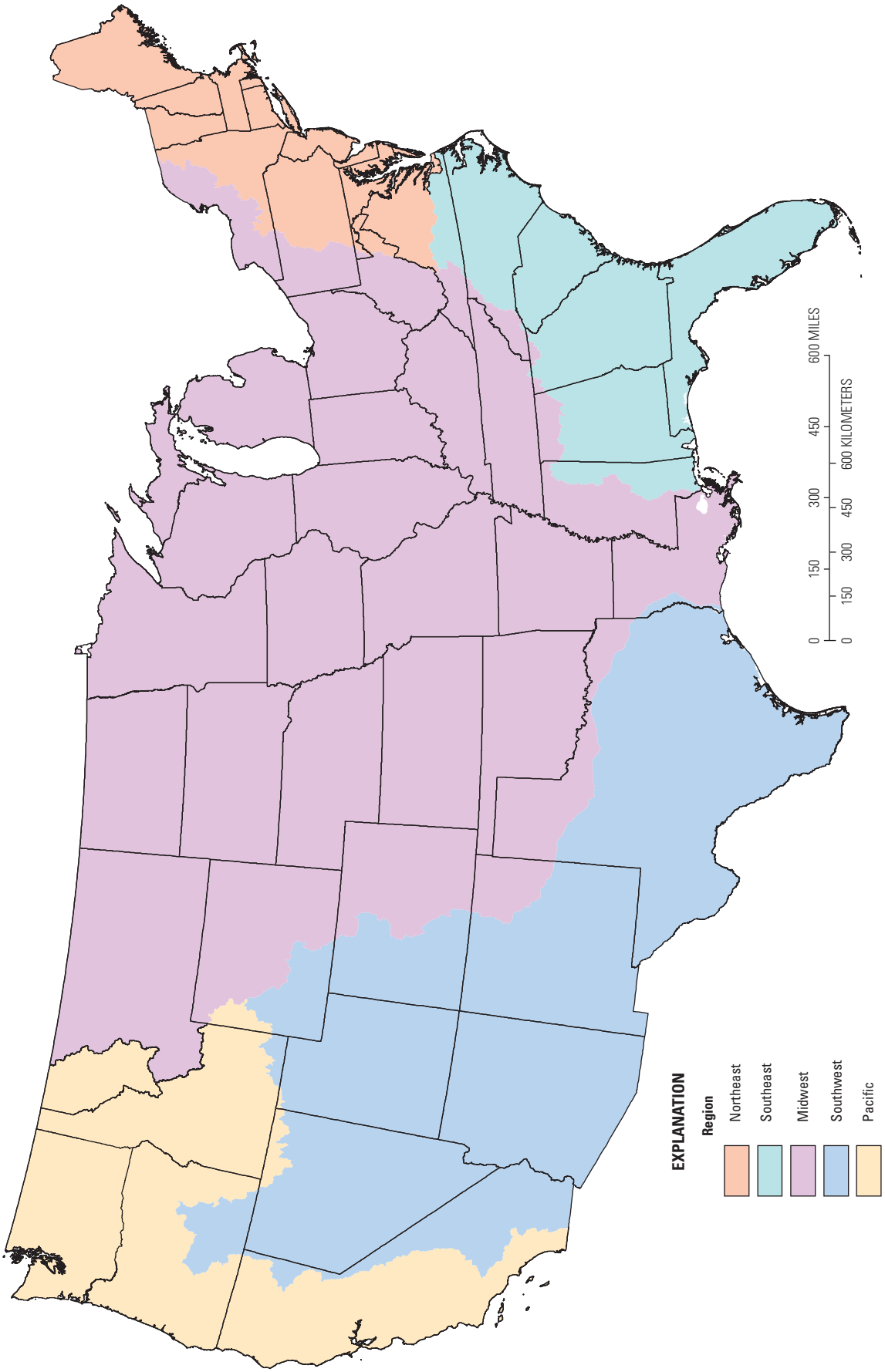


Figure 1. The regions delineated for collection of streamflow and constituent load data for use in Spatially Referenced Regressions on Watershed Attributes (SPARROW) models in the conterminous United States.

monitoring agencies were compiled and evaluated to develop these model objective datasets.

In this report, the detailed approach for assembling, processing, and utilizing available streamflow and water-quality data needed to compute long-term mean daily streamflow and annual load is described. Constituent loads were estimated, and the final set of loads considered for use in the regional SPARROW models were selected from the greater dataset. The distribution of streamflow, TN, TP, and sediment in streams throughout the conterminous United States was estimated from the final data. All streamflow, water-quality data, and load model inputs and outputs are published in Saad and others (2019).

Streamflow and Water-Quality Data Used to Estimate Long-Term Mean Daily Streamflow and Annual Loads

The dependent variable in SPARROW models is long-term mean streamflow, or constituent load normalized (or detrended) to a specific base year (Schwarz and others, 2006; Preston and others, 2009). The base year for the planned regional SPARROW models is 2012; this year was selected so that estimated loads would coincide with available geospatial datasets that describe constituent sources and environmental characteristics. Mean annual constituent loads were estimated using data derived from discrete water-quality samples and continuous (daily) measures of streamflow for each monitoring site.

Streamflow Data

Mean daily streamflow data were compiled for approximately 24,000 stream monitoring sites in the United States. The data from sites that have mean daily streamflow values available from October 1, 1999, to September 30, 2014, were used to compute the mean streamflow volumes and constituent loads considered for use in the SPARROW models. These data were compiled primarily from the U.S. Geological Survey (USGS) National Water Information System (NWIS) database (U.S. Geological Survey, 2019). Additional flow data used in load calculations were obtained from several other sources, including the U.S. Army Corps of Engineers (https://www.usace.army.mil/), the Colorado Division of Water Resources (https://dwr.state.co.us/), and a cooperative streamgaging network operated by the Minnesota Department of Natural Resources (https://www.dnr.state.mn.us/) and the Minnesota Pollution Control Agency (https://www.pca.state.mn.us/) for the Midwest, the International Boundary and Water Commission (https://www.ibwc.gov/home.html) for the Southwest, and the Oregon Water Resources Department (https://www.oregon.gov/OWRD/) for the Pacific.

Water-Quality Data

Previous evaluations of SPARROW models determined that increasing the number of monitoring sites with estimated loads that represent a range in watershed attributes increases the level of certainty in SPARROW model predictions for streams throughout the modeled basin (Schwarz and other, 2006; Preston and others, 2009; Saad and others, 2011). To maximize the number of monitoring sites that could be included in the 2012 regional SPARROW models, an expansive effort was made to inventory, evaluate, and compile discrete water-quality monitoring data from numerous water-resource agencies and organizations, including Federal, State, Tribal, and regional government agencies as well as nongovernmental organizations and universities. The foundations for this national inventory were NWIS (U.S. Geological Survey, 2019) and the U.S. Environmental Protection Agency Storage and Retrieval System (STORET; U.S. Environmental Protection Agency, 2018), which are among the largest national databases of water-quality records available in the United States. All historical water-quality records for streams and rivers publicly housed in the NWIS and STORET databases were initially obtained in March 2013 and September 2013, respectively; however, both datasets were updated to include all available water-quality data from those databases as of March 2015. The updated data were obtained through the Water Quality Portal (http://www.waterqualitydata.us), which is a publicly available access point for all NWIS and STORET water-quality data.

In addition to NWIS and STORET, data were also compiled from numerous regional and local, publicly available water-quality data delivery systems. The effort to inventory water-quality data also included working directly with water resource agencies in each of the 48 contiguous States to acquire relevant digital water-quality records that were not being housed in NWIS, STORET, or other public databases (app. 1).

Nutrient and suspended-sediment data from October 1, 1999, to September 30, 2014, were identified in the inventory and created the subset to be used in SPARROW modeling efforts. For the initial compilation, data from 488 different agencies and organizations were obtained, and the final compiled nutrient and suspended-sediment dataset includes stream water-quality data from 137 different agencies and organizations (app. 1). The initial compilation included more than 30 million water-quality records from approximately 237,000 stream sites. NWIS and STORET contributed more than 90 percent of the nutrient and suspended-sediment records in this dataset; however, efforts to obtain additional data through direct communication with water-resource agencies proved to be useful because that effort filled temporal and spatial data gaps as well as important metadata gaps. For many monitoring sites, filling in these data gaps created a more robust nutrient and suspended-sediment monitoring history from which a monitoring site could be evaluated for use in the regional SPARROW models.

A relatively short list of basic metadata elements was required for a water-quality record to be considered suitable for use in the regional SPARROW modeling efforts. These metadata elements describe site and water-quality sample information and included a station identifier, a station name (which minimally had to include the name of the monitored stream), and locational coordinates. Monitoring sites were indexed to sites in the NHDPlus digital stream network, which was only possible if location coordinates and station name information were available. Required water-quality sample information included the name of the sampled water-quality constituent, sample collection date, numerical value of the analysis, reporting units associated with the numerical value (typically milligrams per liter), remark codes indicating the quality of the measured value or values censored to the laboratory reporting limit, sample fraction (filtered or unfiltered), and weighted units (molecular or elemental—for example, nitrate can be reported as molecular NO_3^- or as elemental nitrogen [N]). If any of these water-quality sample metadata were missing, the sample could not be used in load estimation.

Another important consideration in evaluating monitoring sites for use in the regional SPARROW models was the unification of water-quality records that were collected at the same sampling site by multiple agencies under a single unique identifier. Creating an inventory that was "site-centric" was critical in identifying sites that had long-term monitoring for selected parameters. Unifying water-quality records under a single identifier made it possible to evaluate all the data from a site regardless of any changes to the monitoring agency or the repository that housed the data. In some cases, unifying records for a site under a single identifier resulted in substantial gains in the period of record for a site.

Three techniques were used to unify sites under a single unique identifier: (1) some agencies provided a table that cross-walked their site identifier to USGS or other agency identifiers; (2) some agencies provided cross-walks of site identifiers used by different bureaus within the same agency; (3) station names in the original datasets that included reference to a USGS station were reviewed to identify other agency sampling sites that were colocated with USGS sampling sites or streamgages. When multiple site identifiers needed to be unified under a single identifier, three general rules were used for selecting the final site identifier: (1) the USGS site identifier was selected if available, (2) the most recent site identifier used in a public database was selected, and (3) the site identifier with the longest period of record was selected.

Unifying water-quality records provided an improved dataset to be used for SPARROW modeling. However, not all instances where data could be combined were captured by this initial effort. Many of the sites in each region were investigated more closely, and the data from the sites were combined with additional water-quality records where it was deemed appropriate. Multiple sites on the same stream reach where it was expected that water-quality was not likely to be different were combined and used as a single site. It was assumed that data for sites sampled by different agencies but located on the same stream reach (or to a lesser extent, on nearby stream reaches for the Southwest; fig. 1), with no intervening point sources, no distinct changes in land use, and not on headwater reaches would be suitable for combining. Where multiple sites were on the same stream reach but had intervening point sources, distinct changes in land use, or were on headwater reaches, only the downstream-most site was retained. In the Northeast (fig. 1), sites on the same reach deemed to be on different sampling locations were not combined; however, they were retained for load calculations where appropriate. The data were not pared down for this phase of the study; the decision about which of those sites to keep in the SPARROW models will be done during model estimation.

The data used in nutrient load calculations were the concentrations of unfiltered TN as nitrogen and unfiltered TP as phosphorus, both expressed in units of milligrams per liter. If a TN value was not available, it was calculated as the sum of particulate and dissolved (filtered) forms of measured nitrogen, if those data were available. Where total or particulate and dissolved measurements of nitrogen were unavailable, TN was calculated as the sum of the concentrations of ammonium, organic nitrogen, and nitrite plus nitrate (or just nitrate) where possible. TP was based primarily on TP values reported as phosphorus; however, some TP data were reported as phosphate (PO_4^{3-}) and converted to phosphorus. The specific constituents considered when estimating values for the concentrations of TN and TP and the protocols for combining those concentrations are described in Saad and others (2011, supporting information Data S1).

The constituents used in suspended-sediment load calculations were suspended sediment (SS) and total suspended solids (TSS) concentrations, expressed in milligrams per liter. Both of these constituents are used to quantify concentrations of suspended solid-phase materials in surface water (Gray and others, 2000). The analytical methods to determine concentrations of these two constituents differ. The SS measurement involves evaporating the entire amount of water in a sample, whereas the TSS method analyzes only an aliquot of the water-sediment sample. Paired samples of both constituents show that SS values tend to be increasingly larger than TSS values as the amount of sand-size material increases (Gray and others, 2000). Loads were calculated separately for SS and TSS. If sufficient data for both of these constituents were available at a single monitoring site, loads were calculated separately for each constituent. Most of the sediment loads considered for use in the regional SPARROW models are based on TSS; however, some spatial data gaps exist unless SS loads are included.

Protocol for Screening Streamflow and Water-Quality Sites

After the streamflow and water-quality datasets were compiled, a screening process was performed to identify monitoring sites with sufficient data for computing long-term

mean daily streamflow and annual loads. Screening criteria included minimum requirements for availability of streamflow and water-quality data, the ability to confirm the location of the site and index it to the digital stream network used in the models, and for loads, the ability to associate the water-quality site with a nearby streamgage.

The data requirements for streamgages used for load calculations included that the record must have a minimum length of 10 years of daily streamflow values (with no data gaps) and must include data for 2012 (the base year). For streamgages used to calculate long-term mean streamflows, the daily streamflow data requirements were similar, except that the minimum record length was 13 years. Because the period of record for most of the available streamflow data for International Boundary and Water Commission streamgages in the Southwest ended in 2011, the streamgage requirements described above were altered to require data from 2011 for load calculations and a minimum 12-year record length for calculating long-term mean daily streamflows.

The minimum water-quality data requirements for computing loads at a site included a minimum record length of 3 years, at least 24 samples, and at least 3 samples in each season. For this effort, seasons were defined by groupings of the month in which data were collected (winter, December to February; spring, March to May; summer, June to August; and fall, September to November). The water-quality period of record also had to be within 2 years of the 2012 base year. These minimum requirements were evaluated using the data available at a site from October 1, 1999, to September 30, 2014. Sites meeting the minimum requirements are referred to as "potential load sites." The location of each potential load site was manually checked and verified on the digital stream networks used in the models. If a site location could not be reasonably verified, the site was excluded from further consideration.

Protocol for Matching a Water-Quality Site to a Streamgage

Estimates of load require extended periods of coinciding constituent concentrations and streamflow data. Matching a potential load site to a nearby streamgage initially involved identifying gages that met minimum data requirements for flow and then selecting the streamgage with streamflow characteristics that best represented those at the water-quality monitoring location. Load calculations are ideally done using colocated water-quality and streamflow sites; however, use of a nearby streamgage is a common approach. Where colocation was not possible, the following criteria were used to identify suitable streamgages for a nearby water-quality site:

- overlap between water-quality and streamflow data had to be at least 2 years;
- the ratio of watershed areas between the water-quality site and streamflow site is between 0.75 and 1.33; and
- in general, the water-quality and streamflow site had to be on the same stream.

In the Midwest, Northeast, and Southeast (fig. 1), these additional criteria were also used to match a water-quality site to a streamgage:

- if the watershed area of the water-quality site was less than 260 square kilometers, then the streamgage could be on a nearby stream; and
- the streamgage must be within a reasonable distance (aerial distance less than or equal to 40 kilometers) of the water-quality site to represent similar environmental conditions.

If a potential load site had multiple, nearby, suitable streamgages, priority for selection was given to streamgages with the longest period of data overlap, watershed area ratios closest to 1, and the shortest distance to the water-quality site (Saad and others, 2011).

Methods for Estimating Long-Term Mean Daily Streamflows and Annual Loads

SPARROW models require long-term, mean daily streamflow or annual loads for estimation that do not reflect year-to-year variability or transient spatial patterns in rainfall that occur during any single year. To accomplish this, mean daily streamflow or load is calculated at all sites using a long period of record to limit year-to-year variability at a given site. In some cases, sites have long-term trends in either flow or concentration that could bias the estimation of long-term mean flow or load. In these cases, long-term trends in flow, concentration, or both can be removed statistically so that long-term means can be estimated that only account for year-to-year variability (Schwarz and others, 2006). This process (called "detrending") is usually performed relative to a base year that is consistent with geospatial data (such as land use) that represent that specific year; for the regional SPARROW models, 2012 was the year chosen.

The combination of detrending and calculating a long-term mean provides objective data for the model that are representative of conditions during the base year and are representative of average hydrologic conditions. Average streamflow can vary widely from year to year, so a sufficiently long period of record (20 years or more) is needed to identify long-term trends that are not masked by short-term meteorological variations. For this analysis, the 15-year period of record, used for the streamflow and load estimates described in this report, was considered too short for detrending streamflow. For this reason, streamflows were not detrended when estimating long-term mean daily streamflow, and detrended long-term mean annual loads were computed using only detrended concentrations.

Estimating Long-Term Mean Daily Streamflows

Long-term mean daily streamflow (in cubic feet per second) was computed as the average daily streamflow for the period of record for a streamgage, which ranged from 13 to 15 years. The long-term mean daily streamflow estimates and other associated information for each site are included in Saad and others (2019).

Estimating Long-Term Mean Annual Loads

A variety of methods are available to estimate long-term mean annual loads (in kilograms). Lee and others (2016) evaluated 11 methods for estimating long-term mean annual loads for five constituents, including TN, TP, and SS. Of the methods evaluated by Lee and others (2016), the Beale ratio estimator (BRE; Cochran, 1977) performed the best out of a variety of sampling strategies and constituents. Five-parameter regression with Kalman smoothing (Lee and others, 2016) and weighted regression (Hirsch and others, 2010) methods also did reasonably well. Only two of the programs evaluated, Fluxmaster (Schwarz and others, 2006; Lee and others, 2016) and Weighted Regressions on Time, Discharge, and Season (WRTDS; Hirsch and others, 2010), were capable of estimating detrended, long-term, mean annual loads. A limiting factor with WRTDS is that this method generally requires extensive water-quality data (at least 100 values) for long periods of time (greater than 10 years) to produce accurate estimates of load. Many of the sites used in this effort do not have enough data or do not cover a period long enough to consider using WRTDS. Only one of the programs, Fluxmaster, estimates loads using both the BRE and regression methods from a single dataset. Therefore, loads considered for use in the regional SPARROW models were estimated using the Fluxmaster program. The Fluxmaster program uses a Kalman-smoothing algorithm that differs from the version described in Lee and others (2016). The Kalman-smoothing algorithm used for this effort can handle censored data, whereas the version used in Lee and others (2016) did not (app. 2 and app. 3).

For this effort, long-term mean annual loads for each site and constituent, were estimated using the BRE and Fluxmaster-K five-parameter regression (F5K) algorithms. For sites with a significant trend in load (α=0.05), the F5K load estimates were also detrended. The BRE algorithm was implemented in stratified form as described in Cochran (1977). In this application, as many as eight strata were formed by subdividing the record into two classes based on stream discharge (delineated by the 80th percentile of streamflow) and four seasons (January through March, April through June, July through September, and October through December) after combining all data among years. If the number of samples within any stratum was less than 10, then the strata were collapsed by following the protocol described in Lee and others (2016). The Beale ratio of a given stratum is the ratio of the stratum sample means of load to streamflow (Lee and others, 2016, eq. 1). The estimate of load within a stratum for all days is given by the sum of measured daily loads plus the product of the estimated Beale ratio for the stratum multiplied by the total streamflow for all unsampled days in the stratum. The summation of these estimates across all strata then gives the BRE load estimate for all days in the prediction period. The BRE algorithm, as implemented here, produces nondetrended, long-term, mean annual loads.

When trends were detected, detrended long-term mean annual loads were computed using regression methods developed by Cohn (2005) and implemented in Fluxmaster. The regression-based load estimates are based on a five-parameter water-quality model (eq. 1) that relates the logarithm of concentration at time t (c_t) to the logarithm of daily streamflow (q_t), a decimal time term (T_t), sine and cosine functions of decimal time to account for seasonal variation, and a model residual (e_t):

$$c_t = b_0 + b_q q_t + b_T T_t + b_s \sin(2\pi T_t) + b_c \cos(2\pi T t) + e_t, \quad (1)$$

where

b_0, b_q, b_T, b_s, and b_c are coefficients estimated for each site by the ordinary least squares method, or by the adjusted maximum likelihood method (Cohn, 2005) if any of the c_t measurements are censored. The residuals e_t are assumed to be independent and normally distributed with mean=0 and variance=σ_e^2.

Because of the relatively short (less than 20 years) streamflow record used for load estimation, only concentration data were detrended as part of the F5K load-estimation method. The logarithms of detrended daily concentrations were computed using equation 1 with T_t=2012.5, which represents June 30 of the base year (2012). The logarithms of detrended daily concentrations were then added to the logarithms of daily streamflow to obtain the detrended logarithms of daily load. A serial correlation structure was used to apply the Kalman smooth to the daily load estimates (app. 2). These load estimates were then converted from logarithmic space to real space using methods described by Cohn (2005) and Schwarz and others (2006). For both the BRE and F5K load estimates, the long-term mean annual load was computed by identifying those years included in the analysis period that had no days with missing streamflows, summing the daily load estimates for those years, and dividing by the number of included years to obtain mean annual load in kilograms per year.

Load estimates were evaluated for trend, precision, and bias prior to finalizing selected loads considered for inclusion as model objectives in the regional SPARROW models. In part, the appropriate load estimate was selected for a site based on the presence of a statistically significant trend in load (α=0.05). If there was no significant trend, the BRE long-term mean annual load was selected as the final load. If there was a significant trend in load (p<0.05), the detrended F5K estimate

was used as the final load. The only instance detrended F5K load was not used when the trend was significant was when the water-quality period of record was centered near the base year (anytime within 2012).

Detrending essentially pivots the load estimates for the period of record around the base year (Schwarz and others, 2006, fig. 1.6). If the base year is near the beginning or end of the period of record, the detrended mean annual load can be quite different from the mean annual load. When the period of record is centered near the 2012 base year, the detrended mean annual load should be similar to the mean annual load. In the case of a site with a significant trend but with a period of record that is centered near the base year, the BRE and detrended F5K methods should both produce load estimates that are acceptable for SPARROW modeling. In these cases, the BRE algorithm is preferred because of better overall performance in estimating long-term mean annual loads (Lee and others, 2016).

In addition to trend, load estimates were evaluated for precision (all sites) and bias (for sites with significant trend in load) prior to final selection for consideration as regional SPARROW model objectives. Sites with poor precision and large bias were excluded from consideration for use in regional SPARROW models. The exclusion was iterative, and supervisory oversight of provisional models was used. The standard error of a load model estimate provides a measure of the precision of the load model predictions. Standard errors within 50 percent of the mean load estimate were considered acceptable and are consistent with the level used in previous SPARROW modeling studies (Saad and others, 2011).

Bias provides a description of whether the predicted loads are systematically too high or too low relative to the observed loads. Methods of load estimation done in log space, such as many of the regression methods including F5K, can have biased results when the estimations are converted back into real space because of one or both of the following reasons: (1) heteroscedasticity in the residuals from the model and (2) the statistical relation between streamflow and concentration overestimating or underestimating concentrations for certain streamflow regimes.

One of the underlying assumptions of log-based regression models is that the residuals are normally distributed and have a constant variance (Runkel and others, 2004). If the residuals do not have constant variance (for example, they have a heteroscedastic distribution), then the retransformation process (converting predictions from log space into real space) can lead to biased predictions (Hirsch, 2014).

The computations using the BRE algorithm are done in arithmetic space and will not produce biased results associated with retransformation. In addition to retransformation, there can be a variety of reasons for biased predictions, and the BRE algorithm is generally robust to handle many of them (Lee and others, 2016). For this reason, comparing the F5K load to the BRE load should be an effective way of evaluating bias in the F5K load. Bias was only evaluated for sites with a significant trend in load where the F5K load estimate was being considered as the final load. Bias (R) in a F5K regression estimate, calculated as the BRE load estimate divided by the F5K load estimate, was considered minimal when R is near 1.0; an R between 0.66 and 1.5 was considered acceptable. The selection of this range was chosen to be similar in magnitude to the 50-percent criterion used for model precision. For detrended loads, R can deviate from 1.0 even if the load estimate has no bias. The amount of deviation depends on the magnitude of the trend in load and the proximity of the water-quality period of record compared to the base year. For sites with significant trends, an adjustment to the acceptable bias range was required to account for this deviation. The trend-adjusted acceptable bias range was increased by a factor m, as follows:

$$m = e^{(k \times yrdif)}, \quad (2)$$

where

k is equal to the absolute value of the trend in the load (expressed as a fraction of the mean load estimate per year) and

$yrdif$ is equal to the absolute value of 2012.5 minus the midyear point of the water-quality period of record.

Therefore, R could fall within a range between $m \div 1.5$ and $m \times 1.5$. As the magnitude of the trend gets smaller and the midyear of the water-quality record approaches the base year (2012.5), $m \approx 1.0$. This adjustment assumes that the water-quality samples were approximately equally spaced throughout the sample period, which is not always the case.

For each site and constituent, the following decision steps summarize the process of deciding which load estimate should be considered for inclusion in the regional SPARROW load models:

1. Is the trend in load significant and the water-quality period of record not centered within the 2012 base year?

2. If the answer to step 1 is "no," then evaluate the BRE load estimate for this site. If the standard error for the BRE load estimate is <0.5, then consider including this load in the SPARROW model; if the standard error is ≥0.5, then exclude the load from this site from the SPARROW model objective dataset.

3. If the answer to step 1 is "yes," then evaluate the detrended F5K regression load estimate for this site. If the standard error for the F5K load is <0.5 and the bias is within the acceptable range of $m \div 1.5$ to $m \times 1.5$, then consider including this load in the SPARROW model objective dataset; otherwise, exclude this load.

The final load estimates for TN, TP, and sediment, load model coefficients, and other associated information for each site are included in Saad and others (2019).

Final Streamflow and Load Estimates Considered for Use in the 2012 Regional SPARROW Models

The initial compilation of streamflow data consisted of measurements from approximately 24,000 sites. Only about 21 percent of the streamflow sites evaluated (5,180 of 24,171) passed the screening process (table 1). Sites considered for inclusion in 2012 SPARROW streamflow models are described in Saad and others (2019), along with the estimated long-term mean daily streamflow from those sites.

The initial compilation of suspended-sediment (table 2) and nutrient (table 3) data consisted of more than 30 million water-quality observations from approximately 237,000 sites, collected by 488 sampling agencies. The process of identifying sites that met minimum water-quality screening criteria, could be matched to a streamgage, and produced load estimates that passed evaluation reduced the number of sites considerably. In general, only about 1 to 3 percent of monitored stream sites had enough suspended-sediment or nutrient data and a suitable, nearby streamgage to be considered for load estimation and inclusion in 2012 regional SPARROW models. The final sediment and nutrient load sites that meet minimum water-quality screening criteria accounted for water-quality observations from approximately 3,000 stream sites for sediment and 3,300 stream sites for nutrients, sampled by 137 different agencies and organizations (app. 1). The sites considered for inclusion in the regional 2012 SPARROW sediment and nutrient models are described in Saad and others (2019), along with estimated long-term mean annual loads and associated load model coefficients.

Most of the final load estimates considered for use in the 2012 regional SPARROW nutrient and sediment models were based on the BRE algorithm. Approximately 70, 75, and 82 percent of the TN, TP, and TSS or SS load estimates, respectively, considered for use in the regional models were based on the BRE algorithm. Although the maximum allowable standard error was 50 percent, the average error ranged from 8 percent for TN to 20 percent for TSS and SS; the average standard error for TP was approximately 14 percent. Additionally, although the acceptable range for R was about 0.66 to 1.5, the average, for sites with a significant trend in load, ranged from 1.06 for TN to 1.25 for TSS and SS; the average bias for TP was 1.20. This indicates that BRE load estimates were, on average, higher than detrended F5K load estimates for sites with significant trend in load. This is consistent with the average trend in load, which was negative for all constituents.

Table 1. Daily streamflow sites considered for inclusion in 2012 regional Spatially Referenced Regressions on Watershed Attributes (SPARROW) streamflow models for the conterminous United States.

Model region	Sites with daily streamflow data	Sites considered for inclusion in regional SPARROW streamflow models
Midwest	11,575	2,057
Northeast	2,233	766
Pacific	2,457	817
Southeast	3,121	629
Southwest	4,785	911

Streamflow and Constituent Yields for Sites Considered for Use in the 2012 Regional SPARROW Models

Flow and constituent yields were calculated as mean daily streamflow and annual load, respectively, divided by the watershed area of the monitoring location. Streamflow yields are in millimeters per year (fig. 2). Constituent yields

Table 2. Suspended-sediment water-quality sites considered for inclusion in 2012 regional Spatially Referenced Regressions on Watershed Attributes (SPARROW) sediment models for the conterminous United States.

[SPARROW, Spatially Referenced Regressions on Watershed Attributes; SS, suspended sediment; TSS, total suspended solids; USGS, U.S. Geological Survey]

Model region	Sites with TSS or SS stream water-quality data	Sites with TSS or SS data that meet minimum water-quality screening criteria	Sites with TSS or SS water-quality data matched to a USGS streamgage	Sites with load data considered for inclusion in regional SPARROW sediment models (TSS sites)	Sites with load data considered for inclusion in regional SPARROW sediment models (SS sites)
Midwest	60,360	2,853	1,370	1,085	177
Northeast	15,628	988	535	411	53
Pacific	11,933	741	298	218	53
Southeast	32,016	2,169	896	587	47
Southwest	15,179	1,043	445	329	68

Table 3. Nutrient water-quality sites considered for inclusion in 2012 regional Spatially Referenced Regressions on Watershed Attributes (SPARROW) nutrient models for the conterminous United States.

[SPARROW, Spatially Referenced Regressions on Watershed Attributes; USGS, U.S. Geological Survey; TN, total nitrogen; TP, total phosphorus]

Model region	Sites with nutrient stream water-quality data	Sites with nutrient data that meet minimum water-quality screening criteria	Sites with nutrient water-quality data matched to a USGS streamgage	Sites with load data considered for inclusion in regional SPARROW nutrient models (TN sites)	Sites with load data considered for inclusion in regional SPARROW nutrient models (TP sites)
Midwest	93,994	3,137	1,470	1,344	1,346
Northeast	28,210	875	522	483	279
Pacific	26,889	733	265	140	243
Southeast	47,746	3,418	1,271	834	815
Southwest	24,842	1,128	495	307	428

are in kilograms per square kilometer or metric tons per square kilometer (figs. 3–6). Yield ranges shown on maps represent quantile distributions.

Streamflow Yields

Approximately 5,200 streamflow sites qualified to be considered for inclusion in the 2012 regional SPARROW streamflow models (fig. 2). The distribution of flow sites covers the entire country, with generally higher site density in the eastern parts and lower site density in the western parts. Streamflow yield from these sites range from less than (<) 0.004 to greater than (>) 10,000 millimeters per year (mm/yr). The highest yields are for several large springs with relatively small (and inaccurately defined) watershed areas that were included in the streamflow dataset. The highest yield for a stream or river is approximately 4,400 mm/yr. Extensive areas of high streamflow yield are in the northeastern and northwestern United States. Additional areas of high streamflow yield are scattered throughout the mountain areas of the western and the southeastern United States. Areas of lowest streamflow yield are in the west-central and southwestern parts of the United States.

Total Suspended Solids and Suspended Sediment Yields

Approximately 2,600 TSS and 400 SS sites qualified to be considered for inclusion in 2012 regional SPARROW sediment models (figs. 3 and 4). The distribution of TSS and SS load sites, combined, covers most of the country, with highest site densities in the eastern half. The distribution of each constituent separately shows some large spatial data gaps. Whereas the number of TSS sites are considerably higher than SS sites across the country, there are some large areas with no TSS sites suitable for load estimation. The States of Maine, Montana, New Hampshire, New Mexico, and Wyoming have few to no TSS sites. Outside of the northeastern United States, most States have at least a few SS load sites. TSS yield from these sites ranges from <1 to approximately 400 metric tons per square kilometer (t/km^2); SS yield ranges from <1 to approximately 1,100 t/km^2. The highest sediment yields are in much of the eastern half of the United States and the Pacific Northwest. Additional areas of high sediment yield are scattered over the Dakotas and the southwestern United States. The lowest sediment yields are scattered over the western and north-central United States, Florida, and the coastal areas of Georgia and the Carolinas.

Total Nitrogen and Total Phosphorus Yields

Approximately 3,300 TN and TP load sites qualified to be considered for inclusion in 2012 regional SPARROW nutrient models (figs. 5 and 6). The distribution of TN and TP load sites covers most of the United States; however, site density is highest in the eastern half. There are some noticeable spatial data gaps in the extreme northeastern and southeastern United States; however, similar or larger gaps in western areas are more common. TN yield ranges from <1 to >4,200 kilograms per square kilometer (kg/km^2); TP yield ranges from <1 to >520 kg/km^2. The most extensive area of high TN and TP yields is in the Midwest, where agriculture is the primary land use. Additional areas of high TN yield include the Mid-Atlantic and the Pacific Northwest near the coast. Additional areas of high TP yield include much of the Pacific coastline, the Florida gulf coast, and agricultural areas of the southern and eastern United States. TP yields in Florida are some of the highest in the country and are generally associated with natural phosphorus deposits and related mining operations (Terziotti and others, 2009). Some of the highest TN and TP yields are also in urban areas, where wastewater discharge and urban runoff can produce relatively high loads in relatively small watersheds. The lowest yields for both TN and TP include most of the western and southwestern United States, in areas of low streamflow yield or nonagricultural or nonurban land use.

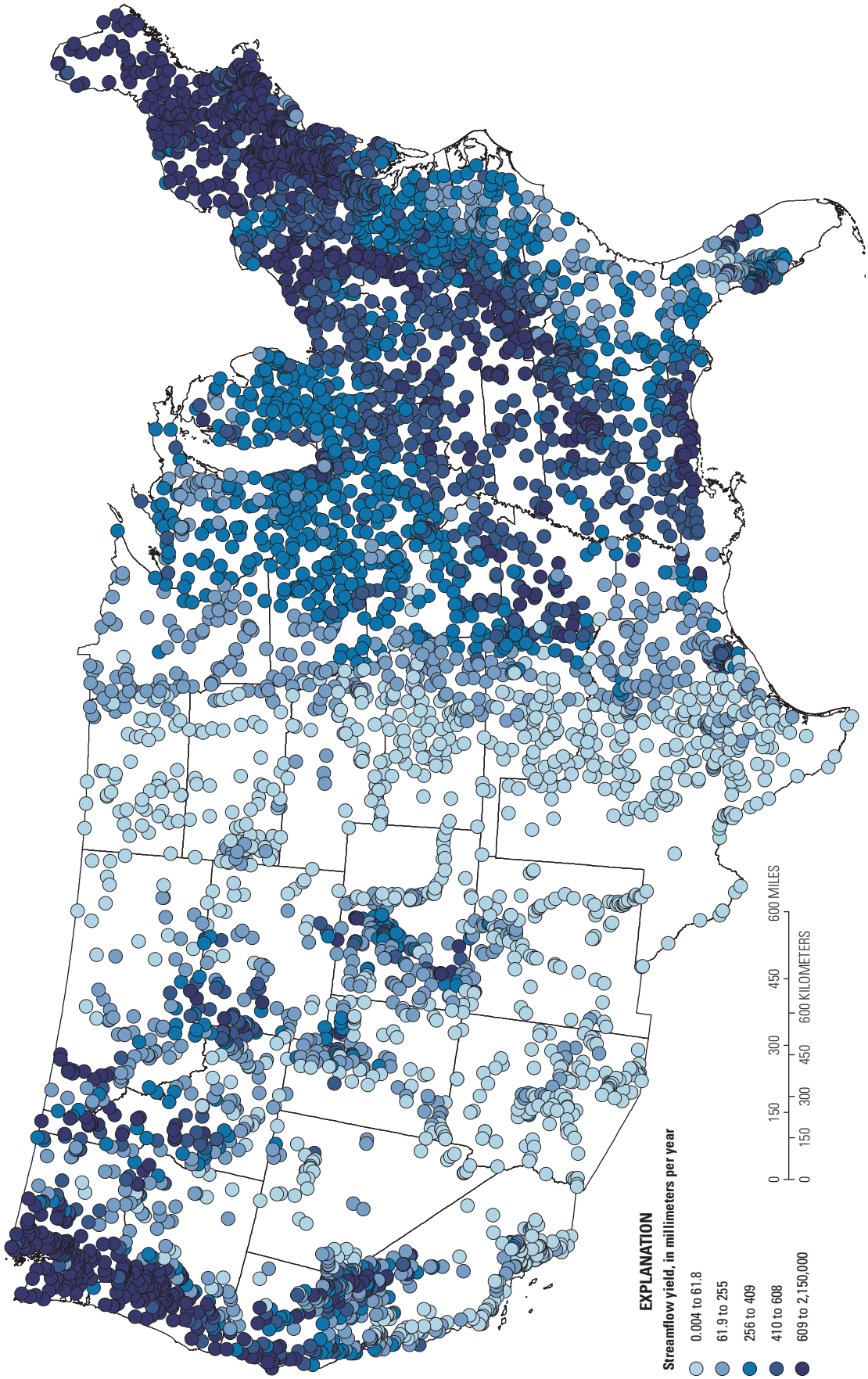


Figure 2. Streamflow yield for stream sites considered for inclusion in 2012 regional Spatially Referenced Regressions on Watershed Attributes (SPARROW) streamflow models.

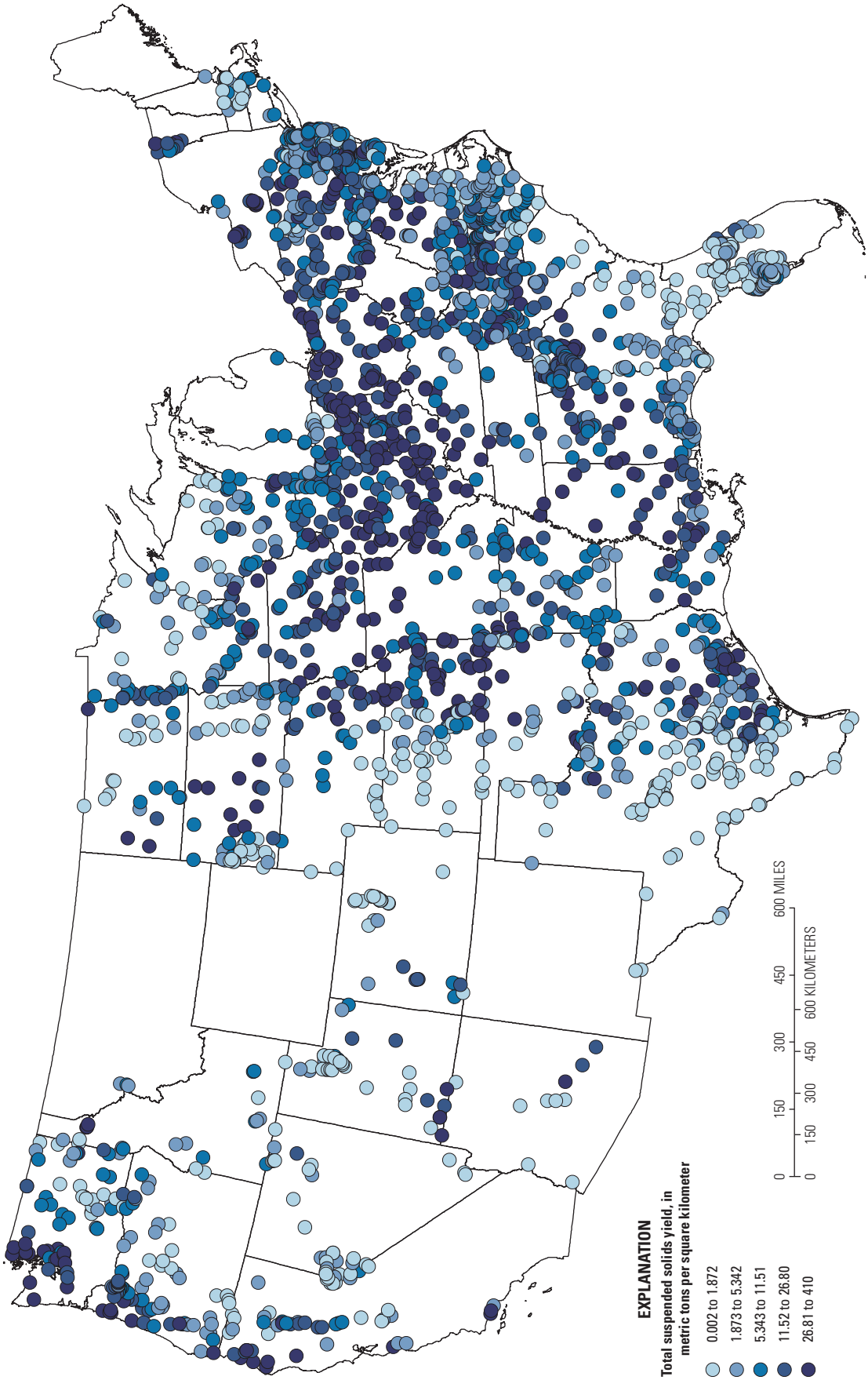


Figure 3. Total suspended solids yield for stream sites considered for inclusion in 2012 regional Spatially Referenced Regressions on Watershed Attributes (SPARROW) sediment models.

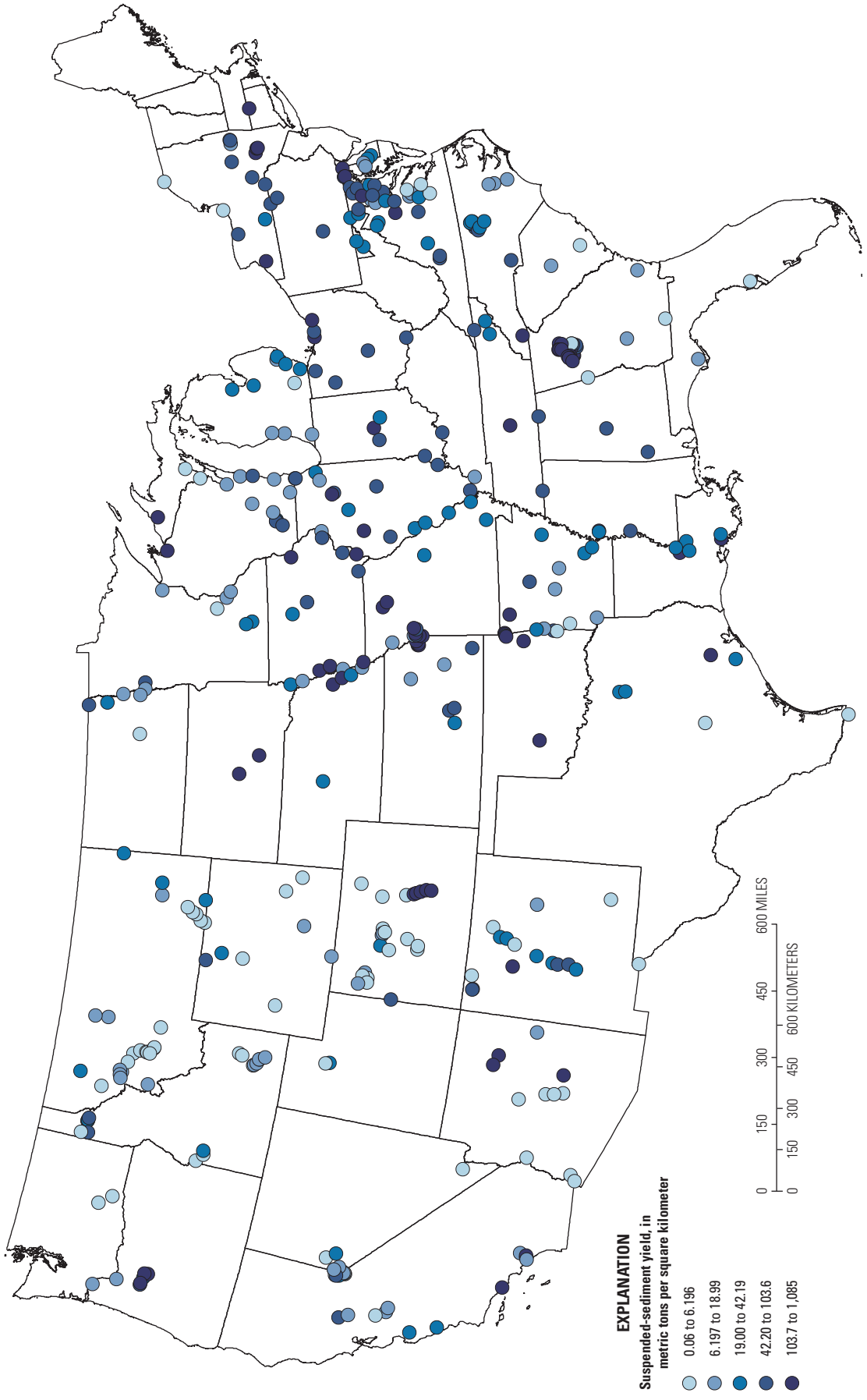


Figure 4. Suspended-sediment yield for stream sites considered for inclusion in 2012 regional Spatially Referenced Regressions on Watershed Attributes (SPARROW) sediment models.

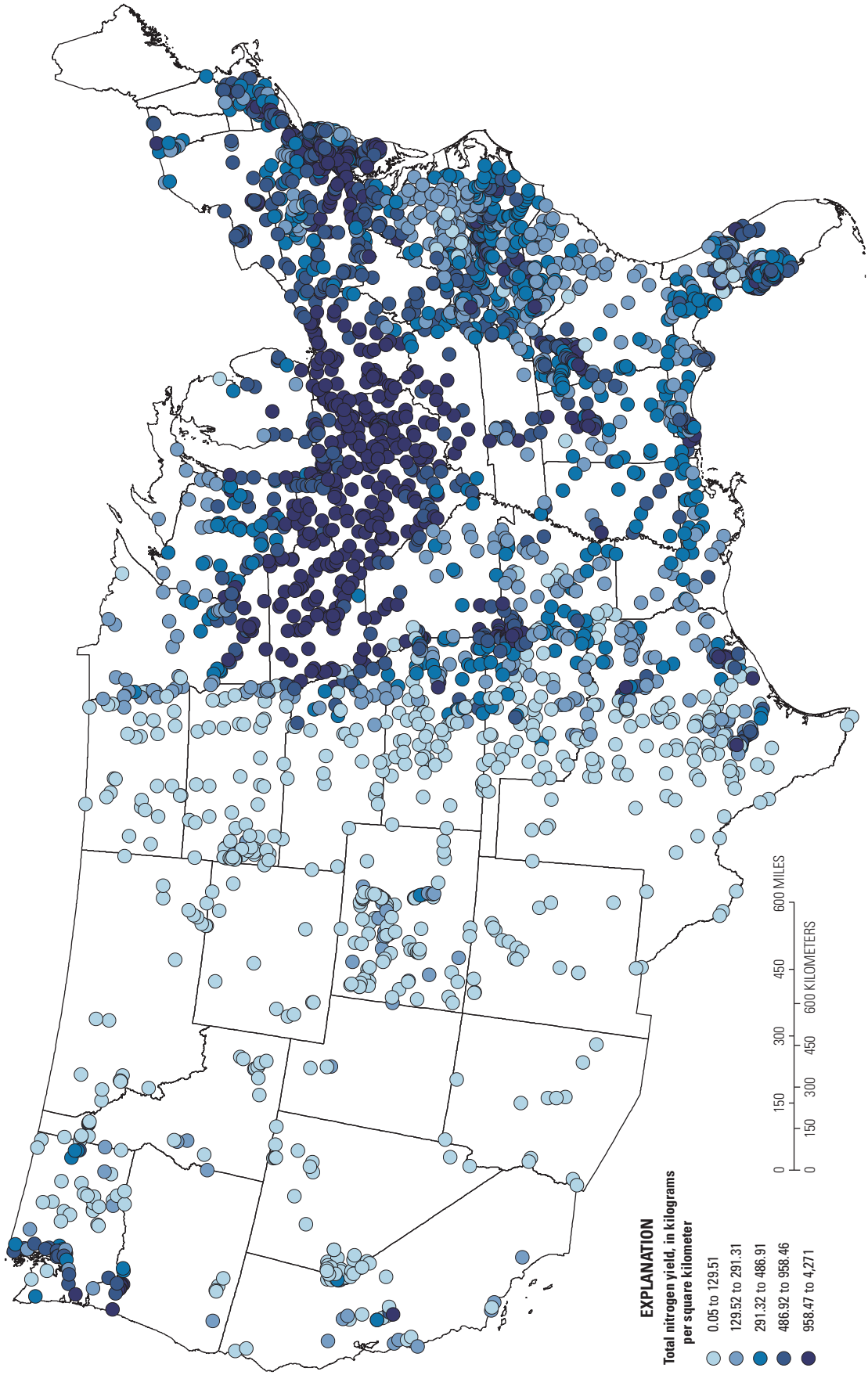


Figure 5. Total nitrogen yield for stream sites considered for inclusion in 2012 regional Spatially Referenced Regressions on Watershed Attributes (SPARROW) nutrient models.

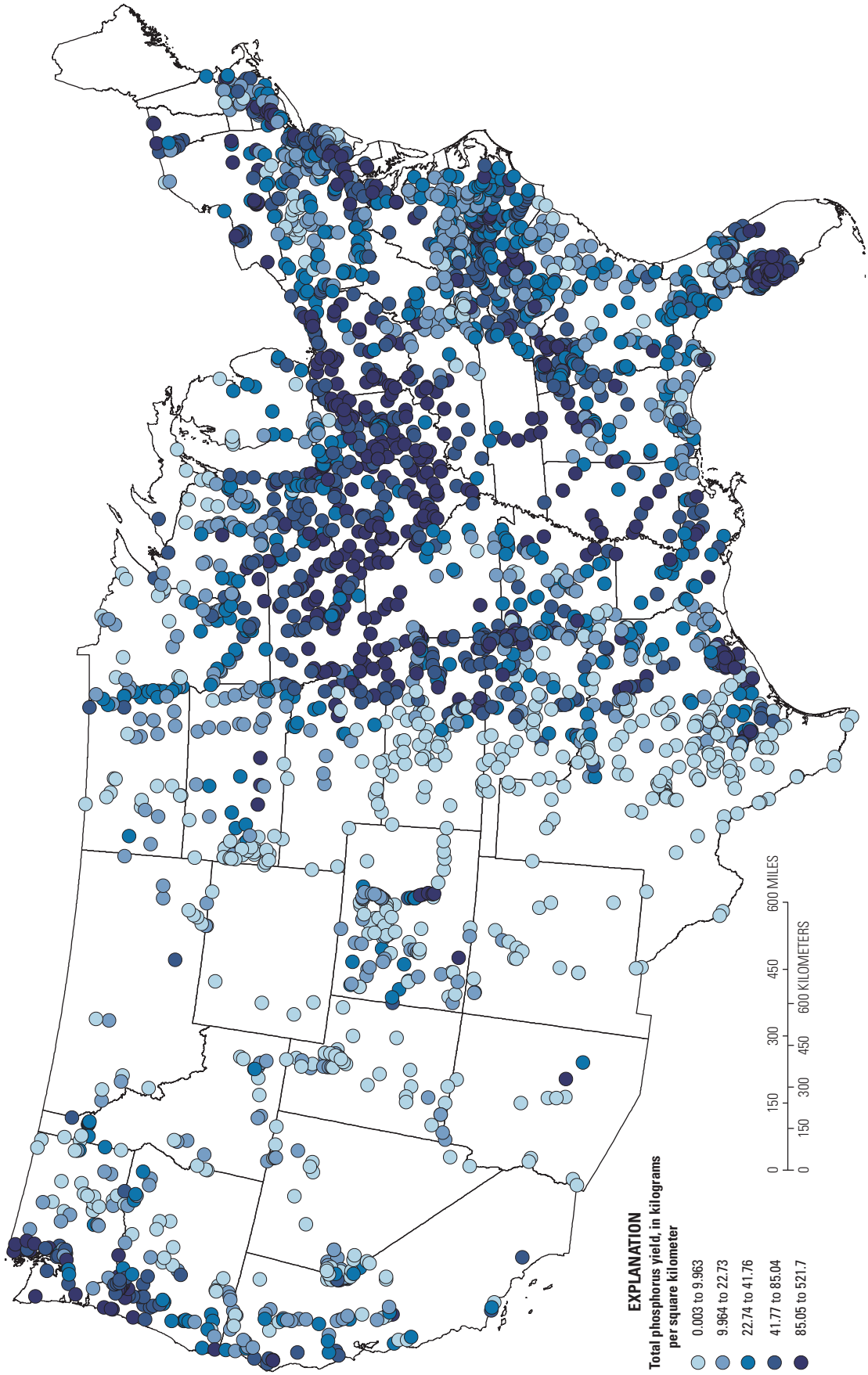


Figure 6. Total phosphorus yield for sites considered for inclusion in 2012 regional Spatially Referenced Regressions on Watershed Attributes (SPARROW) nutrient models.

Summary

Streamflow, nutrient, and sediment concentration data needed to estimate long-term mean daily streamflow and annual constituent loads were compiled from Federal, State, Tribal, and regional agencies, universities, and nongovernmental organizations. The streamflow and loads are planned to be used to develop Spatially Referenced Regressions on Watershed Attributes (SPARROW) models. SPARROW models help describe the distribution, sources, and transport of streamflow, nutrients, and sediment in streams throughout five regions (Midwest, Northeast, Pacific, Southeast, and Southwest) of the conterminous United States. A screening process was applied to reduce nearly 24,000 streamflow sites and more than 237,000 water-quality sites to a select set of approximately 5,200 streamflow, 3,000 sediment, and 3,300 nutrient sites. For these final sites, streamflow and concentration data were of sufficient quality and quantity to be suitable for estimation of long-term mean daily streamflow and annual loads. Estimates from these suitable sites will be considered for use as model objectives in the 2012 (base year) regional SPARROW models. The methods used to estimate long-term mean annual loads include the Beale ratio estimator and Fluxmaster regression method with Kalman smoothing.

The resulting set of qualified streamflow, total nitrogen, total phosphorus, and sediment concentration data represents a significant byproduct of the streamflow- and load-estimation effort—a byproduct that could be useful in other national, regional, or local water-quality assessments. The final streamflow and load sites considered for use in regional SPARROW models represent a wide range in watershed sizes; nitrogen, phosphorus, and sediment sources; and land-use and watershed characteristics in the conterminous United States. The distribution of streamflow and load sites covers most of the United States, but the density of sites is generally greater in the eastern half. The highest streamflow, total nitrogen, total phosphorus, and sediment yields were observed in the humid areas of the Eastern and Pacific Northwest United States. The lowest yields were generally in the west-central and southwestern parts of the country.

References Cited

Bricker, S., Longstaff, B., Dennison, W., Jones, A., Boicourt, K., Wicks, C., and Woerner, J., 2007, Effects of nutrient enrichment in the nation's estuaries—A decade of change: National Oceanic and Atmospheric Administration Coastal Ocean Program Decision Analysis Series 26, 328 p. [Also available at https://doi.org/10.1016/j.hal.2008.08.028.]

Carpenter, S.R., Caraco, N.F., Correll, D.L., Howarth, R.W., Sharpley, A.N., and Smith, V.H., 1998, Nonpoint pollution of surface waters with phosphorus and nitrogen: Ecological Applications, v. 8, no. 3, p. 559–568. [Also available at https://doi.org/10.1890/1051-0761(1998)008[0559:NPOSWW]2.0.CO;2.]

Cochran, W.G., 1977, Sampling techniques (3d ed.): New York, Wiley, 428 p.

Cohn, T.A., 2005, Estimating contaminant loads in rivers—An application of adjusted maximum likelihood to type 1 censored data: Water Resources Research, v. 41, no. 7, W07003. [Also available at https://doi.org/10.1029/2004WR003833.]

Gray, J.R., Glysson, G.D., Turcios, L.M., and Schwarz, G.E., 2000, Comparability of suspended-sediment concentration and total suspended solids data: U.S. Geological Survey Water-Resources Investigations Report 2000–4191, 14 p. [Also available at https://doi.org/10.3133/wri004191.]

Heimann, D.C., Sprague, L.A., and Blevins, D.W., 2011, Trends in suspended-sediment loads and concentrations in the Mississippi River Basin, 1950–2009: U.S. Geological Survey Scientific Investigations Report 2011–5200, 33 p. [Also available at https://doi.org/10.3133/sir20115200.]

Hirsch, R.M., 2014, Large biases in regression-based constituent flux estimates—Causes and diagnostic tools: Journal of the American Water Resources Association, v. 50, no. 6, p. 1401–1424. [Also available at https://doi.org/10.1111/jawr.12195.]

Hirsch, R.M., Moyer, D.L., and Archfield, S.A., 2010, Weighted regressions on time, discharge, and season (WRTDS), with an application to Chesapeake Bay River inputs: Journal of the American Water Resources Association, v. 46, no. 5, p. 857–880. [Also available at https://doi.org/10.1111/j.1752-1688.2010.00482.x.]

Howarth, R., Billen, G., Swaney, D., Townsend, A., Jaworski, N., Lajtha, K., Downing, J., Elmgren, R., Caraco, N., Jordan, T., Berendse, F., Freney, J., Kudeyarov, V., Murdoch, P., and Zhao-Liang, Z., 1996, Regional nitrogen budgets and riverine N & P fluxes for the drainages to the North Atlantic Ocean—Natural and human influences: Biogeochemistry, v. 35, no. 1, p. 75–139. [Also available at https://doi.org/10.1007/BF02179825.]

Howarth, R.W., Sharpley, A., and Walker, D., 2002, Sources of nutrient pollution to coastal waters in the United States—Implications for achieving coastal water quality goals: Estuaries, v. 25, no. 4, p. 656–676. [Also available at https://doi.org/10.1007/BF02804898.]

Lee, C.J., and Glysson, G.D., 2013, Compilation, quality control, analysis, and summary of discrete suspended-sediment and ancillary data in the United States, 1901–2010: U.S. Geological Survey Data Series 776, 35 p., accessed April 4, 2019, at https://doi.org/10.3133/ds776.

Lee, C.J., Hirsch, R.M., Schwarz, G.E., Holtschlag, D.J., Preston, S.D., Crawford, C.G., and Vecchia, A.V., 2016, An evaluation of methods for estimating decadal stream loads: Journal of Hydrology (Amsterdam), v. 542, p. 185–203, accessed April 4, 2019, at https://doi.org/10.1016/j.jhydrol.2016.08.059.

Michalak, A.M., Anderson, E.J., Beletsky, D., Boland, S., Bosch, N.S., Bridgeman, T.B., Chaffin, J.D., Cho, K., Confesor, R., Daloglu, I., DePinto, J.V., Evans, M.A., Fahnenstiel, G.L., He, L., Ho, J.C., Jenkins, L., Johengen, T.H., Kuo, K.C., LaPorte, E., Liu, X., McWilliams, M.R., Moore, M.R., Posselt, D.J., Richards, R.P., Scavia, D., Steiner, A.L., Verhamme, E., Wright, D.M., and Zagorski, M.A., 2013, Record-setting algal bloom in Lake Erie caused by agricultural and meteorological trends consistent with expected future conditions: Proceedings of the National Academy of Sciences of the United States of America, v. 110, no. 16, p. 6448–6452, accessed April 4, 2019, at https://doi.org/10.1073/pnas.1216006110.

Moore, R.B., and Dewald, T.G., 2016, The road to NHDPlus—Advancements in digital stream networks and associated catchments: Journal of the American Water Resources Association, v. 52, no. 4, p. 890–900. [Also available at https://doi.org/10.1111/1752-1688.12389.]

Phillips, S.W., ed., 2002, The U.S. Geological Survey and the Chesapeake Bay—The role of science in environmental restoration: U.S. Geological Survey Circular 1220, 32 p. [Also available at https://doi.org/10.3133/cir1220.]

Preston, S.D., Alexander, R.B., Woodside, M.D., and Hamilton, P.A., 2009, SPARROW modeling—Enhancing understanding of the nation's water quality: U.S. Geological Survey, Fact Sheet 2009–3019, 6 p. [Also available at https://doi.org/10.3133/fs20093019.]

Runkel, R.L., Crawford, C.G., and Cohn, T.A., 2004, Load estimator (LOADEST)—A FORTRAN program for estimating constituent loads in streams and rivers: U.S. Geological Survey Techniques and Methods, book 4, chap. A5, 69 p. [Also available at https://doi.org/10.3133/tm4A5.]

Saad, D.A., Argue, D.M., Schwarz, G.E., Anning, D.W., Ator, S.W., Hoos, A.B., Preston, S.D., Robertson, D.M. and Wise, D.A., 2019, Water-quality and streamflow datasets used for estimating long-term mean daily streamflow and annual loads to be used in regional nutrient, suspended sediment and streamflow SPARROW models, 1999–2014: U.S. Geological Survey data release, accessed April 4, 2019, at https://doi.org/10.5066/F7DN436B.

Saad, D.A., Schwarz, G.E., Robertson, D.M., and Booth, N.L., 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, v. 47, no. 5, p. 933–949. [Also available at https://doi.org/10.1111/j.1752-1688.2011.00575.x.]

Schwarz, G.E., Hoos, A.B., Alexander, R.B., Smith, R.A., 2006, The SPARROW surface water-quality model—Theory, applications and user documentation: U.S. Geological Survey Techniques and Methods, book 6, chap. B3, 248 p. [Also available at https://doi.org/10.3133/tm6B3.]

Smith, R.A., Schwarz, G.E., and Alexander, R.B., 1997, Regional interpretation of water-quality monitoring data: Water Resources Research, v. 33, no. 12, p. 2781–2798. [Also available at https://doi.org/10.1029/97WR02171.]

Terziotti, S., Hoos, A.B., Harned, D.A., and Garcia, A.M., 2009, Mapping watershed potential to contribute phosphorus from geologic materials to receiving streams, southeastern United States: U.S. Geological Survey Scientific Investigations Map 3102, 1 sheet. [Also available at https://doi.org/10.3133/sim3102.]

U.S. Environmental Protection Agency, 1998, National strategy for the development of regional nutrient criteria: U.S. Environmental Protection Agency EPA–822–R–98–002, 47 p. [Also available at https://www.waterboards.ca.gov/sanfranciscobay/water_issues/programs/planningtmdls/amendments/estuarineNNE/EPA%201998%20National%20Strategy%20Document.pdf.]

U.S. Environmental Protection Agency, 2000a, Nutrient criteria technical guidance manual—Lakes and reservoirs: U.S. Environmental Protection Agency EPA–822–B–00–001, 238 p. [Also available at https://www.epa.gov/sites/production/files/2018-10/documents/nutrient-criteria-manual-lakes-reservoirs.pdf.]

U.S. Environmental Protection Agency, 2000b, The quality of our nation's waters—A summary of the national water quality inventory—1998 report to Congress: U.S. Environmental Protection Agency EPA 841–S–00–001, 19 p. [Also available at https://www.epa.gov/sites/production/files/2015-09/documents/2000_07_07_305b_98report_98brochure.pdf.]

U.S. Environmental Protection Agency, 2007, Hypoxia in the northern Gulf of Mexico—An update by the EPA Science Advisory Board: U.S. Environmental Protection Agency EPA–SAB–08–003, 333 p. [Also available at https://yosemite.epa.gov/sab/sabproduct.nsf/C3D2F27094E03F90852573B800601D93/$File/EPA-SAB-08-003complete.unsigned.pdf.]

U.S. Environmental Protection Agency, 2018, Storage and retrieval and water quality exchange: U.S. Environmental Protection Agency Water Quality Portal, accessed April 4, 2019, at https://www.epa.gov/waterdata/storage-and-retrieval-and-water-quality-exchange.

U.S. Geological Survey, 2019, USGS water data for the nation: U.S. Geological Survey National Water Information System database, accessed April 4, 2019, at https://doi.org/10.5066/F7P55KJN.

Appendix 1. Sampling Agencies Associated with Water-Quality Data Used To Compute Mean Annual Load Estimates Considered for Use in 2012 Regional SPARROW Models

Table 1.1. Agencies that provided data to be used to compute mean annual load estimates considered for use in the 2012 regional Spatially Referenced Regressions on Watershed Attributes (SPARROW) models.

Abbreviation	Agency name
AL DEM	Alabama Department of Environmental Management
AR DEQ	Arkansas Department of Environmental Quality
AZ DEQ	Arizona Department of Environmental Quality
AZ TCOCOPAH	Cocopah Indian Tribe Environmental Protection Office, Arizona
CA CEDEN	California Environmental Data Exchange Network
CA DWR	California Department of Water Resources
CA ESJWQC	East San Joaquin Water Quality Coalition, California
CA RWB	California Environmental Protection Agency, Regional Water Boards
CA SJCDWQC	San Joaquin County and Delta Water Quality Coalition, California
CA TYUROKTEP	Yurok Tribe, California
CA WSSJRWC	Westside San Joaquin River Watershed Coalition, California
CO CBS	CBS Operations Incorporated, Colorado
CO CRWWN	The Rivers of Colorado Water Watch Network
CO CSUTL	Colorado Springs Utilities, Colorado
CO CWSD	Centennial Water and Sanitation District, Colorado
CO DPHE	Colorado Department of Public Health and Environment
CO MWRD	Metropolitan Waste Water Reclamation District, Colorado
CO SACWSD	South Adams County Water and Sanitation District, Colorado
CO SWQC	Summit Water Quality Committee, Colorado
CO THORN	City of Thornton, Colorado
CT DEEP	Connecticut Department of Energy and Environmental Protection
DC DDOE	District of Columbia Department of the Environment
DE DGS	Delaware Geological Survey
DE DNREC	Delaware Department of Natural Resources and Environmental Control
FL BCEPD	Broward County Environmental Protection Department, Florida
FL BCSUD	Brevard County Stormwater Utility Department, Florida
FL CCC	City of Cape Coral, Florida
FL CCPC	Collier County Pollution Control, Florida
FL CPG	City of Punta Gorda, Florida
FL DEP	Florida Department of Environmental Protection

Abbreviation	Agency name
FL HLCE	Hillsborough County Environmental, Florida
FL LCPW	Leon County Public Works, Florida
FL LCRM	Lake County Water Resource Management, Florida
FL LRD	Loxahatchee River District, Florida
FL LW	Florida Lakewatch
FL MCEMD	Manatee County Environmental Management Department, Florida
FL MCGL	McGlynn Laboratories Incorporated, Florida
FL NWFWMD	Northwest Florida Water Management District
FL OCEP	Orange County Environmental Protection, Florida
FL PCDEM	Pinellas County Department of Environmental Management, Florida
FL SEMC	Seminole County, Florida
FL SFWMD	South Florida Water Management District
FL SJWMD	Saint John's Water Management District, Florida
FL SRWMD	Suwannee River Water Management District, Florida
FL SWFWMD	Southwest Florida Water Management District
FL VCEHL	Volusia County Environmental Health Lab, Florida
GA DNR	Georgia Department of Natural Resources
IA DMWW	Des Moines Water Works, Iowa
IA DNR	Iowa Department of Natural Resources
IA GSWSP	Iowa Geological Survey Watershed Snapshots
IA ISUCOE	Iowa State University, Des Moines River Water Quality Network
IA VWMP	Iowa Volunteer Water Monitoring Program
ID DEQ	Idaho Department of Environmental Quality
IL EPA	Illinois Environmental Protection Agency
IL MWRDGC	Metropolitan Water Reclamation District of Greater Chicago, Illinois
IL SWS	Illinois State Water Survey
IN CEG	Indianapolis Citizen Energy Group, Indiana
IN DEM	Indiana Department of Environmental Management
IN EPWU	City of Elkhart Public Works and Utilities, Indiana
IN MBWQ	Muncie Bureau of Water Quality, Indiana
IN WREC	Wabash River Enhancement Corporation, Indiana
KS DHE	Kansas Department of Health and Environment
KY DEP	Kentucky Department for Natural Resources and Environmental Protection
LA DEQ	Louisiana Department of Environmental Quality
MA DEP	Massachusetts Department of Environmental Protection
MA MRWA	Massachusetts Rural Water Association
MD DNR	Maryland Department of Natural Resources
MD DOE	Maryland Department of the Environment
MI DEQ	Michigan Department of Environmental Quality
MI TLRBOI	Little River Band of Ottawa Indians, Michigan
MN DA	Minnesota Department of Agriculture
MN MCES	Metropolitan Council Environmental Services, Minnesota
MN PCA	Minnesota Pollution Control Agency

Abbreviation	Agency name
MN TRLDNR	Red Lake Division of Natural Resources, Minnesota
MO DNR	Missouri Department of Natural Resources
MO MACT	Mactec Incorporated, Missouri
MO MEC	Midwest Environmental Consultants, Missouri
MO VRSR	Versar Incorporated, Missouri
MS DEQ	Mississippi Department of Environmental Quality
NC DENR	North Carolina Department of Environment and Natural Resources
NC TEBCI	Eastern Band of Cherokee Indians, North Carolina
ND DOH	North Dakota Department of Health
NE DEQ	Nebraska Department of Environmental Quality
NH DES	New Hampshire Department of Environmental Services
NJ BTMUA	Brick Township, New Jersey Municipal Utilities Authority
NJ DEP	New Jersey Department of Environmental Protection
NJ HDG	New Jersey Harbor Discharges Group
NJ MCHD	Monmouth County, New Jersey Health Department
NJ RCEWRP	Rutgers University Cooperative Extension, New Jersey
NM ISC LRG COMP	New Mexico Interstate Stream Commission Lower Rio Grande Compendium
NM TPJ	Pueblo of Jemez, New Mexico
NM TPT	Pueblo of Taos, New Mexico
NV DECNR	Nevada Department of Conservation and Natural Resources
NV SNWA	Southern Nevada Water Authority (Environmental Monitoring and Management)
NY COMSI	Community Science Institute, New York
NY DEC	New York Department of Environmental Conservation
OH EPA	Ohio Environmental Protection Agency
OH HDLBG	Heidelberg University, National Center for Water Quality Research, Ohio
OH MCD	Miami Conservancy District, Ohio
OH NEORSD	Northeast Ohio Regional Sewer District
OK CONC	Oklahoma Conservation Commission
OK DEQ	Oklahoma Department of Environmental Quality
OK TKN	Kaw Nation of Oklahoma
OK TSFN	Sac and Fox Nation, Oklahoma
OK TWN	Wyandotte Nation, Oklahoma
OK WRB	Oklahoma Water Resources Board
OR DEQ	Oregon Department of Environmental Quality
OR KTRS	Klamath Tribes Natural Resources Department, Oregon
OR PWB	Portland Bureau of Environmental Services, Oregon
PA DEP	Pennsylvania Department of Environmental Protection
RG3 SJRWI	Saint Joseph River Watershed Initiative
RG3 SRBC	Susquehanna River Basin Commission
RG3 TSWQC	Tri-State Water Quality Council
RG4 DRBC	Delaware River Basin Commission

Abbreviation	Agency name
RG8 ORSANCO	Ohio River Valley Water Sanitation Commission (ORSANCO)
SC DHEC	South Carolina Department of Health and Environmental Control
SD DENR	South Dakota Department of Environment and Natural Resources
TN DEC	Tennessee Department of Environment and Conservation
TX CEQ	Texas Commission on Environmental Quality
TX MCWE	Meadows Center for Water and the Environment, Texas
US ACE	U.S. Army Corps of Engineers
US BR	Bureau of Reclamation
US DA	U.S. Department of Agriculture
US EPA	U.S. Environmental Protection Agency
USGS	U.S. Geological Survey
US NPS	National Park Service
UT DEQ	Utah Department of Environmental Quality
UT USU	Utah State University
VA DEQ	Virginia Department of Environmental Quality
VT DEC	Vermont Department of Environmental Conservation
WA SDE	Washington State Department of Ecology
WA TK NRD	Kalispel Tribe Natural Resources Department, Washington
WI DNR	Wisconsin Department of Natural Resources
WI GBMSD	Green Bay Metropolitan Sewerage District, Wisconsin
WI THCN	Ho-Chunk Nation, Wisconsin
WI TSMC	Stockbridge-Munsee Community, Wisconsin
WV DEP	West Virginia Department of Environmental Protection

Appendix 2: A Kalman-Smoothing Estimate of Water-Quality Loads Based on Simulated Maximum Likelihood Estimation for Censored Data: the Fluxmaster-K Algorithm

This appendix describes a statistical method for estimating a water-quality model of the type described by Cohn (2005) with the added feature that residuals of the model can be serially correlated according to a first-order autoregressive (AR(1)) process. The complication arising from such a model is that water-quality measurements are often censored, causing substantial complication for maximum likelihood methods commonly used to estimate the model. The Simulated Maximum Likelihood Estimation (SMLE) method (Lee, 1999) provides a viable approach to estimation. Various properties of the estimates are established using concepts developed in Schendler (2005).

The utility of the AR(1) water-quality model is that predictions of load can be improved using Kalman smoothing methods whereby load predictions can be conditioned directly on the water-quality measurements (Harvey, 1989). For models without serially correlated residuals, such conditioning results in simply replacing the prediction of load for a day containing an observation with the monitored value, all other non-monitored days being unaffected by the conditioning information; however, with serially correlated residuals the conditioning implies predictions for intervening days between observations can be optimally smoothed to improve the precision of the predictions. The complication of incorporating censored information into these predictions can be assuaged using the same simulation methods employed in model estimation. The method is further extended to obtain a conditioned estimate of detrended mean annual load. The uncertainty of the estimate is evaluated using a parametric bootstrap approach, as described in Schwarz (2006). The load estimation algorithm presented here extends to censored data the previously described Fluxmaster-K method for uncensored data appearing in Lee and others (2016).

Model Estimation

Consider a model of daily water quality in which the logarithm of contaminant concentration for day t, denoted y_t, is linearly related to a $1 \times K$ row vector of predictor variables, X_t, a $K \times 1$ vector of associated coefficients, β, and a residual error term, v_t, according to

$$y_t = X_t\beta + v_t. \tag{2.1}$$

The error term in (1) is assumed to follow an AR(1) process, such that

$$v_t = \rho v_{t-1} + \varepsilon_t, \tag{2.2}$$

where ρ is the daily serial correlation coefficient, with $0 \leq \rho \leq 1$, and ε_t is an independent, normally distributed random innovation to the random v_t process, with mean equal to 0 and variance, σ_ε^2.

Additional structure is added to the model to address the fact that water quality is typically measured infrequently. Let the water-quality sample consist of N measurements, ordered over time and sequentially indexed by $i = 1 \ldots N$, and let d_i refer to the number of days between

successive measurements $i-1$ and i, so that $d_i = t_i - t_{i-1}$, where t_i refers to the time period for observation i. Equations 2.1 and 2.2 can then be combined to yield

$$y_{t_i} = X_{t_i}\beta + \rho^{d_i}\left(y_{t_{i-1}} - X_{t_{i-1}}\beta\right) + e_{t_i}, \tag{2.3}$$

where equation 2.2 is recursively applied d_i times to account for the gap in days between observations i and $i-1$, this recursion leading to the d_i exponent on ρ, v_{t_i} and $v_{t_{i-1}}$ being replaced using equation 2.1 evaluated at t equal to t_i and t_{i-1}, and e_{t_i} corresponds to the exponentially weighted sum of innovations, ε_t, over the period $t = t_{i-1}+1$ to $t = t_i$, causing e_{t_i} to be an independent (across i), normally distributed random error having zero mean and variance $\sigma_i^2 = \sigma_\varepsilon^2(1-\rho^{2d_i})/(1-\rho^2)$. Note that for the first observation of the sample, d_i is infinity, implying $\rho^{d_1} = 0$, and $\sigma_1^2 = \sigma_\varepsilon^2/(1-\rho^2)$.

It is further assumed that water quality is observed only through measurement, and that measurements may be censored. Let the measurement of the logarithm of concentration for observation *i* be denoted ω_i, where $\omega_i = y_{t_i}$ if y_{t_i} exceeds the censoring threshold for that observation, denoted $\bar{y}_{t_i}$; otherwise, if $y_{t_i} < \bar{y}_{t_i}$, the measurement is censored and $\omega_i = \bar{y}_{t_i}$. In the sequel the subscripted time notation is dropped and observations are referred to solely by their order through time, so that $y_i \equiv y_{t_i}$.

The existence of censoring implies maximum likelihood is an appropriate methodology for the estimation of the $(K+2)\times 1$ vector of model parameters, $\theta = \{\beta', \rho, \sigma_\varepsilon\}'$. However, the presence of serially correlated errors makes this a non-standard estimation problem. To understand the complication, consider the log-likelihood for the N measurements comprising the sample, $\Omega_N = \{\omega_i : i = 1, \ldots, N\}$, which can be expressed via Bayes rule (Amemiya, 1985) as the sum of *N* conditional log likelihoods

$$L \equiv \ln \Pr(\Omega_N|\theta) = \sum_{i=1}^{N} \ln \Pr(\omega_i|\Omega_{i-1}, \theta), \tag{2.4}$$

where $\Omega_i = \{\omega_j : j = 1, \ldots, i\}$, and $\Pr_i(\omega_i|\Omega_{i-1}, \theta)$ is the probability of obtaining measurement ω_i conditioned on the parameters and measurements up through observation $i-1$. If observation i is uncensored, so that $\omega_i = y_i$, then $\Pr_i(\omega_i|\Omega_{i-1}, \theta)$ represents a probability density function; otherwise, if observation i is censored, with $\omega_i = \bar{y}_i$, then $\Pr_i(\omega_i|\Omega_{i-1}, \theta)$ is a cumulative probability of observing the event $y_i \le \bar{y}_i$. Without serial correlation the conditional probabilities could be written as $\Pr_i(\omega_i|\Omega_{i-1}, \theta) = \Pr_i(\omega_i|\theta)$, which has no dependence on past measurements. Conversely, with no censoring, the AR(1) process implies the conditional probability could be restricted to $\Pr_i(\omega_i|\Omega_{i-1}, \theta) = \Pr_i(\omega_i|\omega_{i-1}, \theta)$, and the previous measurement would then suffice for all the information from the past.

A more complicated situation arises if observations are serial correlated and some observations are censored. In this case the censored measurement is no longer sufficient for all past information, and the conditioning information pertinent to observation i includes all past measurements back to the most recent previously observed uncensored measurement, or to the beginning of the sample, whichever comes first. Thus, if the most recent previous uncensored measurement pertains to observation $i-h$, then the conditional probability for observation i is given by

$$\Pr_i(\omega_i|\Omega_{i-1}, \theta) = \frac{\Pr(\omega_i, \ldots, \omega_{i-h+1}|y_{i-h}, \theta)}{\Pr(\omega_{i-1}, \ldots, \omega_{i-h+1}|y_{i-h}, \theta)} =$$

$$\frac{\int_{-\infty}^{\bar{y}_{i-h+1}} \cdots \int_{-\infty}^{\bar{y}_{i-1}} \Pr_i(\omega_i|y_{i-1}^*,\theta) p_{i-1}(y_{i-1}^*|y_{i-2}^*,\theta) \cdots p_{i-h+1}(y_{i-h+1}^*|y_{i-h},\theta) dy_{i-1}^* \cdots dy_{i-h+1}^*}{\int_{-\infty}^{\bar{y}_{i-h+1}} \cdots \int_{-\infty}^{\bar{y}_{i-1}} p_{i-1}(y_{i-1}^*|y_{i-2}^*,\theta) \cdots p_{i-h+1}(y_{i-h+1}^*|y_{i-h},\theta) dy_{i-1}^* \cdots dy_{i-h+1}^*}, \quad (2.5)$$

where $p_j(a|b)$ is the marginal probability of variable a conditioned on information b in period j, and the y^* variables in the last term are the surrogates for the y variables that are integrated over their feasible range given by the upper censoring limit. Note that with equation 2.5 inserted into equation 2.4, and if successive observations $i-1$ and i are censored, then the logarithm of the numerator of observation $i-1$ cancels with the logarithm of the denominator of observation i. Consequently, all censored observations, except possibly for the last observation of the sample if it is censored, are dropped from the log likelihood in equation 2.4, leaving only the logarithms of the numerators of equation 2.5 for which the observation is uncensored (except for the last observation in the sample if it is censored). Because the numerical stability of the algorithm is improved with probabilities expressed as in equation 2.5, terms for all observations, both censored and uncensored, are retained in computing the log likelihood.

The maximum dimension of the integrals appearing in equation 2.5, across all observations, depends on the maximum number of successive censored observations in the sample. Standard numerical methods can be used for evaluating integrals up to three dimensions (Press and others, 1986). For integrals of greater dimension, however, a feasible approach is integration by a Monte Carlo simulation. Advances in these methods have greatly improved the accuracy and efficiency with which large dimensioned integrals are evaluated.

In published econometrics literature, the problem of parameter estimation with serial correlation and censored data has been studied by Lee (1999) using a variant of the Gibbs sampler (Robert and Casella, 1999) to generate simulated values of the unobserved censored observations, with the random samples selected efficiently to conform with censoring thresholds. The evaluation of a likelihood function via this approach is SMLE.

The basic idea of the simulation method is to randomly draw S sequences, indexed by $s = 1, \ldots, S$, of the dependent variable for all observations that are censored. The dependent variables are drawn from conditional probability distributions that depend only on the random draw of the dependent variable for the previous observation for the same sequence, *s*. Lee (1999) proposed imposing an additional restriction that the distribution generating the random values conform with the known censoring threshold for the observation $\bar{y}_i$. In this way, all simulated values of the dependent variable are acceptable in terms of being consistent with the actual observed data. In the vernacular of Monte Carlo simulation, the Gibbs sampler is a variant of the more general Metropolis-Hastings algorithm in which the acceptance probability of the simulated value equals one. Generally, a higher acceptance probability improves the numerical efficiency of the simulation algorithm.

The simulated conditional probability for the case where the previous observation is censored is given by

$$\widehat{\Pr}_i(\omega_i|\Omega_{i-1},\theta) = S^{-1} \sum_{s=1}^{S} \Pr_i\left(\omega_i \middle| \tilde{y}_{i-1}^{(s)}, \theta\right) w_{i-1}^{(s)}, \quad (2.6)$$

where $\tilde{y}_{i-1}^{(s)}$ is a simulated value of the dependent variable, randomly drawn from a conditional distribution (described in the next paragraph), and $w_{i-1}^{(s)}$ is an appropriately selected weight that unbiases the sampling method.

A critical element of the simulation algorithm concerns the generation of the simulated $\tilde{y}_j^{(s)}$, for some censored observation j. As discussed above, Lee (1999) advocates a Gibbs sampler that imposes the information that observation being simulated is censored at $\bar{y}_j$ and is correlated with the previous simulated value. Consequently, $\tilde{y}_j^{(s)}$ is drawn from a truncated distribution with probability density

$$\tilde{y}_j^{(s)} \sim p_j\left(\tilde{y}\middle|\tilde{y}_{j-1}^{(s)},\theta\right)/P_j\left(\bar{y}_j\middle|\tilde{y}_{j-1}^{(s)},\theta\right), \tag{2.7}$$

where $P_j\left(\bar{y}_j\middle|\tilde{y}_{j-1}^{(s)},\theta\right)$ corresponds to the cumulative probability for the event $\tilde{y} \le \bar{y}_j$, conditioned dynamically on the previous value $\tilde{y}_{j-1}^{(s)}$, and normalizes the distribution to have unit probability over the range $\tilde{y} \le \bar{y}_j$. Given the assumptions of the innovation term in equation 2.2, the distribution in equation 2.7 is truncated normal, a distribution from which random values are readily generated, as shown below in equation 2.13.

The expectation of equation 2.6 given the sampling scheme described in equation 2.7 is given by

$$E\left[S^{-1}\sum_{s=1}^{S}\Pr_i\left(\omega_i\middle|\tilde{y}_{i-1}^{(s)},\theta\right)w_{i-1}^{(s)}\right] =$$

$$\frac{\int_{-\infty}^{\bar{y}_{i-h+1}}\cdots\int_{-\infty}^{\bar{y}_{i-1}}\Pr_i(\omega_i|y_{i-1}^*,\theta)p_{i-1}(y_{i-1}^*|y_{i-2}^*,\theta)\cdots p_{i-h+1}(y_{i-h+1}^*|y_{i-h},\theta)w_{i-1}^{(s)}dy_{i-1}^*\cdots dy_{i-h+1}^*}{P_{i-1}(\bar{y}_{i-1}|y_{i-2}^*,\theta)\cdots P_{i-1}(\bar{y}_{i-h+2}|y_{i-h+1}^*,\theta)P_{i-h+1}(\bar{y}_{i-h+1}|y_{i-h},\theta)}. \tag{2.8}$$

The sampling scheme results in an unbiased estimate of equation 2.5 if

$$w_{i-1}^{(s)} = \frac{P_{i-1}\left(\bar{y}_{i-1}\middle|\tilde{y}_{i-2}^{(s)},\theta\right)\cdots P_{i-h+2}\left(\bar{y}_{i-h+2}\middle|\tilde{y}_{i-h+1}^{(s)},\theta\right)P_{i-h+1}(\bar{y}_{i-h+1}|y_{i-h},\theta)}{\int_{-\infty}^{\bar{y}_{i-h+1}}\cdots\int_{-\infty}^{\bar{y}_{i-1}}p_{i-1}(y_{i-1}^*|y_{i-2}^*,\theta)\cdots p_{i-h+1}(y_{i-h+1}^*|y_{i-h},\theta)dy_{i-1}^*\cdots dy_{i-h+1}^*}. \tag{2.9}$$

The denominator of equation 2.9 is a multi-dimensioned integral, complicating its evaluation. Simulation methods using the same sampling scheme in equation 2.7 can be used to approximate the denominator by computing the average,

$$S^{-1}\sum_{s=1}^{S}P_{i-1}\left(\bar{y}_{i-1}\middle|\tilde{y}_{i-2}^{(s)},\theta\right)\cdots P_{i-h+2}\left(\bar{y}_{i-h+2}\middle|\tilde{y}_{i-h+1}^{(s)},\theta\right)P_{i-h+1}(\bar{y}_{i-h+1}|y_{i-h},\theta), \tag{2.10}$$

which, using the derivation of equation 2.8 as a guide, has an expectation that can be shown to equal the denominator of equation 2.9. Under general conditions, the limit of equation 2.10 as S approaches infinity equals its expectation (Amemiya, 1985), implying that in the limit the weights will exactly generate the theoretical likelihood.

Thus, the simulated weights, denoted $\widehat{w}_i^{(s)}$, take the form

$$\widehat{w}_i^{(s)} = \frac{P_i\left(\bar{y}_i\middle|\tilde{y}_{i-1}^{(s)},\theta\right)\cdots P_{i-h+2}\left(\bar{y}_{i-h+2}\middle|\tilde{y}_{i-h+1}^{(s)},\theta\right)P_{i-h+1}(\bar{y}_{i-h+1}|y_{i-h},\theta)}{S^{-1}\sum_{s=1}^{S}P_i\left(\bar{y}_i\middle|\tilde{y}_{i-1}^{(s)},\theta\right)\cdots P_{i-h+2}\left(\bar{y}_{i-h+2}\middle|\tilde{y}_{i-h+1}^{(s)},\theta\right)P_{i-h+1}(\bar{y}_{i-h+1}|y_{i-h},\theta)}. \tag{2.11}$$

The weights in equation 2.11 have an average of 1 across all simulations, a property that improves numerical stability in evaluating the likelihood function (Lee, 1999). Moreover, the last term in both the numerator and denominator of equation 2.11 is $P_{i-h+1}(\bar{y}_{i-h+1}|y_{i-h},\theta)$, which does not

vary with simulation sequence and therefore can be canceled from the ratio. Consequently, if $h = 1$, implying observation $i - 1$ is uncensored, then the only term included in either the numerator or denominator is the last term, making $\widehat{w}_i^{(s)} = 1$ for all S simulation sequences. In addition, if observation $i - h + 1$ references the first observation in the sample (that is, $i = h = 1$), then there is no prior conditioning information and $P_{i-h+1}(\bar{y}_{i-h+1}|y_{i-h}, \theta) = P_{i-h+1}(\bar{y}_{i-h+1}|\theta)$ continues to be a constant across simulation sequences. In this way, the weights can accommodate any sampling outcome, including cases in which the first observation is either censored or uncensored.

The estimated weights, described by equation 2.11 and subsequently inserted into equation 2.6, can be constructed recursively according to the relation

$$\widehat{w}_j^{(s)} = \begin{cases} \dfrac{P_j\left(\bar{y}_j \middle| \tilde{y}_{j-1}^{(s)}, \theta\right)}{\widehat{\Pr}_j\left(\omega_j \middle| \Omega_{j-1}, \theta\right)} \widehat{w}_{j-1}^{(s)} & \text{if } i - h + 1 < j \leq i - 1, \\ 1 & \text{if } j = i - h + 1. \end{cases} \tag{2.12}$$

The weight for the case where $j = i - h + 1$ is one because the conditional distribution for y_{i-h+1}^* is solely a function of y_{i-h}, which is uncensored and thus requires no simulation to evaluate the conditional likelihood. As observations are processed sequentially, each time the observation is uncensored the weights are re-initialized to $\widehat{w}_j^{(s)} = 1$, a consequence of conditioning for the AR(1) process in which the most recent uncensored observation is sufficient for all prior information.

The estimation of the coefficient vector θ requires iterative methods, implying the full likelihood given in equation 2.4 must be repeatedly simulated for each iteration. To assure stability across iterations, it is necessary that changes in the likelihood be attributed to changes in the underlying coefficients rather than from random variations arising from different draws of the random sequences. To achieve this stability, a single $N \times S$ matrix U of random variables, with elements $u_i^{(s)}$, is independently drawn from the uniform [0,1] distribution. Given that the underlying errors of the water-quality model are normally distributed, the conditional sequence $\tilde{y}_j^{(s)}$ for an observation censored at $\bar{y}_j$ is generated according to a truncated normal random number generating function

$$\tilde{y}_j^{(s)} = z_j^{(s)} + \sigma_j \Phi^{-1}\left(u_j^{(s)} \Phi\left(\left(\bar{y}_j - z_j^{(s)}\right)/\sigma_j\right)\right), s = 1, \ldots S, \tag{2.13}$$

where $z_j^{(s)} = X_j\beta + \rho^{d_j}\left(\tilde{y}_{j-1}^{(s)} - X_{j-1}\beta\right)$, $\Phi(\cdot)$ is the standard normal cumulative distribution function, and $\Phi^{-1}(\cdot)$ is the inverse standard normal (Lee, 1999).

The algorithm for estimating the log likelihood (eq. 2.4) using simulation methods is detailed in the "Algorithm for the Evaluation of the Log-Likelihood Function Using the Simulation Method" section of this appendix. The determination of the parameter vector θ that maximizes the simulated log likelihood is referred to as SMLE. The algorithm is designed such that if the sample includes no censored observations, then the method reverts to standard maximum likelihood.

The likelihood function defined by equation 2.5 is not a standard likelihood function in that integrating it over its support $y_i \in (-\infty, \infty)$ rather than equaling 1, as with a standard likelihood function, is not defined due to the presence of the cumulative distribution term if observation i is censored, a function that does not vary with y_i causing the integral to be infinite (Amemiya, 1973,

Davidson and MacKinnon, 1993). Consequently, asymptotic properties of the standard maximum likelihood estimators cannot be assumed. Amemiya (1973) has verified the standard maximum likelihood estimator asymptotic properties of consistency, asymptotic normality, and asymptotic efficiency for the Tobit model, the model of independent censored values. Schnedler (2005) has extended Amemiya's result to include models in which the observations consist of a vector of censored values, the values within the vector being dependent but the vectors across observations being independent. This result would suffice for the present analysis if a record of N concentration values was interpreted as a single observation on a vector of N dependent values and asymptotic samples were assumed to consist of independent draws of N-value records; however, this is not quite consistent with the assumed data-generation process in which extensions of the record cannot be assumed to be independent.

Gallant and White (1988) established large-sample properties of consistency and asymptotic normality for the coefficients of non-linear optimization problems with serially dependent observations. The requirements for these results are general, restricting the dependence between observations to meet mixing and near epoch dependence (NED) conditions, and placing regularity conditions on the objective function. The homoscedastic Gaussian AR(1) data-generating process assumed for ν_t (the residual in eq. 2.1) conforms with strong mixing (Gallant and White, 1988). If the random components of the X_t explanatory data, which is principally streamflow, are a stationary Gaussian autoregressive moving average (ARMA) process, then these components are also strong mixing (Gallant and White, 1988).

Near epoch dependence is the condition that the log-likelihood for any given observation has a dependence on past values of ν_t and X_t that diminishes at a sufficient rate, a condition that is trivially established if the log-likelihoods for all observations depend on only a finite number of lagged values of these variables (Gallant and White, 1988). Under NED conditions, a function of a mixing process effectively behaves as a mixing process. If the variability of the observation log-likelihood is sufficiently bounded, a condition that limits the ability to incorporate trend into the analysis, the log-likelihood can be shown to converge pointwise to its expected value. An additional constraint on the continuity of the observation log-likelihood ensures convergence is uniform in the model coefficients, implying the coefficient estimates converge to the true values (consistency), almost surely.

Andrews (1992) provides simpler conditions that assure consistency in probability, a weaker convergence condition than 'almost surely,' but acceptable for establishing asymptotic properties. As explained by Davidson (1994), the mixing and NED conditions from Gallant and White (1988) that restrict dependence are retained, but the additional conditions suggested by Andrews (1992) can be met by a generally valid restriction on the partial derivatives of the objective function with respect to the model coefficients. The same derivative conditions can be used to establish asymptotic normality (Andrews, 1992). Andrews and McDermott (1995) show that consistency and asymptotic normality are retained even if the model includes deterministically trending variables.

A complication in applying these conditions for the present analysis is that an observation of the log likelihood is not strictly defined. As explained above, the log likelihood given in equation 2.4 is essentially the sum of log-transformed values of the numerator of equation 2.5 for uncensored observations. This term consists of lagged values of the mixing variables, but the number of lags depends on the number of successive censored observations that precede the uncensored observation, which is the outcome of a random process. Progress can be made by adopting the approach of Schnedler (2005), which defines "visibility states" based on the pattern

of uncensored or censored observations observed for a given sample (see appendix 3 for a description of visibility states and their application in proving regularity conditions for the log-likelihood function and efficiency properties of the coefficient estimates). Under this approach it is necessary to assume that certain visibility states occur with zero probability; for example, if the number of uncensored observations in an infinite sample is finite, then it would not be possible for coefficients to converge on their true values. Such a situation could occur if there was a downward trend in concentration with a fixed censoring threshold. If such states are excluded, the remaining states are those for which an infinite sample contains an infinite number of uncensored observations, and the number of lags for any given observation is necessarily finite. It would then be necessary to show that the conditions of Gallant and White (1988) and Andrews (1992) hold for any arbitrary visibility state in which there are an infinite number of uncensored observations in a sample of infinite observations.

A formal proof of the consistency and asymptotic normality properties is beyond the scope of this report. For the purposes of the discussion in this appendix, the dependency conditions proposed by Gallant and White (1988), along with the differentiability conditions described by Davidson (1994) are assumed to be met. Such an assumption is justified in this case because it is straightforward to show that these conditions are met if there are an infinite number of uncensored observations and all censored observations are simply eliminated from the sample.

If the conditions proposed by Gallant and White (1988) or Davidson (1994) are met, then the coefficients maximizing the log-likelihood function described by equations 2.4 and 2.5, denoted $\hat{\theta}$, are consistent, meaning that the estimates go to the true values θ_0 in probability as sample size N approaches infinity, and that they have an asymptotic normal distribution,

$$\lim_{N\to\infty} \sqrt{N}\left(\hat{\theta} - \theta_0\right) \sim N(0, A^{-1}BA^{-1}), \tag{2.14}$$

where A is the $(K+2)\times(K+2)$ matrix $A = \lim_{N\to\infty} E[N^{-1}L_{\theta\theta'}(\Omega,\theta_0)]$, assumed to be nonsingular to permit the existence of the inverse matrix A^{-1}, and B is the $(K+2)\times(K+2)$ matrix $B = \lim_{N\to\infty} E[N^{-1}L_{\theta}(\Omega,\theta_0)L_{\theta'}(\Omega,\theta_0)]$; the $(K+2)\times 1$ vector $L_{\theta}(\Omega,\theta_0)$ and the $(K+2)\times(K+2)$ matrix $L_{\theta\theta'}(\Omega,\theta_0)$ are the first- and second-order partial derivatives of the log-likelihood function given by equations 2.4 and 2.5 evaluated at the true values of the parameters. Continuity and identifiability conditions that imply θ_0 is an unconstrained maximum of the expectation of the log-likelihood function imply that the maximum occurs at a concave section of the expected log-likelihood, making A negative definite.

Schendler (2005) shows that, for any given vector of correlated observations, under standard regularity conditions for the probability density function, the expectation of the second-order partial derivative of the log-likelihood for that vector equals the negative of the expectation of the outer product of the first-order partial derivative with its transpose, where all derivatives are evaluated at θ_0; (see appendix 3 for a description of Schendler's analysis). Applying this result to the present analysis, with the vector interpreted as the entire record of N observations, then $E[L_{\theta\theta'}(\Omega,\theta_0)] = -E[L_{\theta}(\Omega,\theta_0)L_{\theta'}(\Omega,\theta_0)]$, implying $A = -B$ in the description of the asymptotic covariance matrix in equation 2.14, and the covariance can be expressed equivalently as $-A^{-1}$ or B^{-1}. Thus, using the latter form the asymptotic covariance matrix is given by

$$\lim_{N\to\infty} E\left[\left(\hat{\theta} - \theta_0\right)\left(\hat{\theta} - \theta_0\right)'\right] \equiv AV\left[\hat{\theta}\right] = \lim_{N\to\infty} E[N^{-1}L_{\theta}(\Omega,\theta_0)L_{\theta'}(\Omega,\theta_0)]\,. \tag{2.15}$$

Methods developed by Schendler (2005) can be used to show the asymptotic covariance matrix in equation 2.15 achieves the Cramer-Rao lower bound (see app. 3) implying that, assuming the

conditions specified by Gallant and White (1988) or Davidson (1994) are met, the maximization of the log-likelihood function described by equations 2.4 and 2.5 gives estimates that are asymptotically efficient.

Cohn (2005) recommends an adjustment based on work of Shenton and Bowman (1977) that removes first-order bias in the coefficient estimates (the expected difference between the estimated coefficients and the true coefficients if the difference is multiplied by the number of observations and the sample size goes to infinity). Shenton and Bowman (1977) show that the first-order bias of coefficient $\hat{\theta}_k$, denoted B_k, can be expressed as

$$E\left[\hat{\theta}_k - \theta_k\right] \equiv B_k \cong \sum_{a=1}^{K+2} V_{ka} \sum_{b=1}^{K+2} \sum_{c=1}^{K+2} E\left[L_{ab}L_c + \frac{1}{2}L_{abc}\right] V_{bc}$$

$$= \sum_{a=1}^{K+2} V_{ka} \sum_{b=1}^{K+2} \sum_{c=1}^{K+2} E\left[L_{ab}L_c - \frac{1}{2}(L_{ab}L_c + L_{ac}L_b + L_{bc}L_a + L_aL_bL_c)\right] V_{bc}$$

$$= -\frac{1}{2} \sum_{a=1}^{K+2} V_{ka} \sum_{b=1}^{K+2} \sum_{c=1}^{K+2} E[L_{bc}L_a + L_aL_bL_c] V_{bc}$$

$$= -\frac{1}{2} V_{k\cdot} E[L_\theta \mathrm{vec}(L_{\theta\theta'} + L_\theta L_{\theta'})'] \mathrm{vec}(V), \qquad (2.16)$$

where V_{ka} is the covariance between estimated coefficients $\hat{\theta}_k$ and $\hat{\theta}_a$ – the element in the kth row and ath column of the coefficient covariance matrix $V\left[\hat{\theta}\right]$ (see eq. 2.15), $V_{k\cdot}$ is the kth row of the covariance matrix, L_c is the partial derivative of the log-likelihood function with respect to coefficient θ_c ($L_c \equiv \partial L/\partial\theta_c$, evaluated at $\theta = \theta_0$), L_{ab} is the second-order partial deriviative of the log-likelihood function with respect to coefficients θ_a and θ_b, ($L_{ab} \equiv \partial^2 L/\partial\theta_a\partial\theta_b$, evaluated at $\theta = \theta_0$), L_{abc} is the third-order partial derivative of the log-likelihood function ($L_{abc} \equiv \partial^3 L/\partial\theta_a\partial\theta_b\partial\theta_c$, evaluated at $\theta = \theta_0$), and vec(·) converts a square $(K+2) \times (K+2)$ matrix into a $(K+2)^2 \times 1$ vector by stacking the columns in left to right order.

The second equality in equation 2.16 expands the terms of L_{abc}, eliminating the third-order partial derivative term through an argument similar to the one used to show $A = -B$ in the derivation of the covariance matrix (see app. 3); this simplifies L_{abc} to consist only of first- and second-order partial derivatives of the log-likelihood function. The third equality follows by combining the $L_{ab}L_c$ terms and then noting that these cancel with the $L_{ac}L_b$ term because the covariance matrix is symmetric, implying $E[L_{ab}L_c]V_{bc} = E[L_{ab}L_c]V_{cb}$, and upon switching labels b and c, which are arbitrary in summation, yields $E[L_{ab}L_c]V_{bc} = E[L_{ac}L_b]V_{bc}$. The last equality follows because the product $[L_{bc} + L_bL_c]V_{bc}$ summed over all b and c indices is equivalent to a simple inner product of the vectorized square matrices $L_{\theta\theta'} + L_\theta L_{\theta'}$ and V. The expectation operator forms expectations for the variance log-likelihood differential terms under alternative realizations of the model innovations, assuming that the censoring threshold for each observation day is known, with the log-likelihood being formulated to be consistent with the censoring threshold applied under alternative outcomes for the innovations.

The evaluation of the expectation terms appearing in equation 2.15 for the determination of the covariance matrix and in equation 2.16 for the determination of the first-order bias adjustment are obtained using simulation methods, where S sequences of N standard-normal distributed random variables are independently drawn to generate simulated sequences of the dependent variable based on the SMLE estimated values of the parameters. Censoring thresholds for each observation are then applied to simulate sequences of measurements. Because censoring thresholds for uncensored observations are not generally known, the median of the censoring thresholds among the censored observations is used as the censoring threshold for the uncensored observation days. Given each simulation sequence of measurements, the log-likelihood of that sequence is evaluated using the algorithm described in the "Algorithm for the Evaluation of the Log-Likelihood Function Using the Simulation Method" section of this appendix, and the first- and second-order partial derivatives of the log-likelihood function with respect to the parameter vector, as required in equations 2.15 and 2.16, are evaluated using the numerical central-difference approximation method (Conte and de Boor, 1972). In the case of sample records without any censored observations, a censoring threshold is not known, and the simulation of measurements does not apply any censoring limits to the simulated dependent variable. Consequently, the evaluation of the likelihood is much simpler with each of the N terms of the log-likelihood being statistically independent.

The first-order partial derivatives for each simulation sequence are substituted into equation 2.15 and averaged, the inverse of this average being the estimate of the covariance matrix. Bias is obtained using these covariances for the two covariance terms appearing in equation 2.16. To estimate the expectation term in equation 2.16, simulation averages of the first- and second-order partial derivatives of the log-likelihood function are substituted for the relevant terms in the brackets of the expectation term, the evaluation of which is then averaged to approximate the expectation term.

The bias-adjusted coefficients are evaluated to ensure that the coefficients adhere to estimation bounds: the adjusted ρ must be between 0 and 0.999, and the adjusted σ is required to be greater than 0.01. The adjustment is also evaluated to ensure it does not induce too large a change in the log-likelihood function, where the adjustment is rejected if the log-likelihood evaluated at the maximum-likelihood coefficient estimates is greater than the log-likelihood evaluated at the adjusted estimates by an amount greater than the 0.999999 quantile of the chi-square distribution with $K + 2$ degrees of freedom. If the adjusted coefficients violate any of these conditions then the model is re-estimated with the serial correlation coefficient constrained to equal 0 using a method that approximates the method developed by Cohn (2005), as described in Schwarz and others (2006). The serial correlation parameter is also set to zero if the median number of days between observations in the water-quality record is greater than 30; large gaps in the typical water-quality record make it difficult to obtain a reliable estimate of the serial correlation parameter. In these cases, the Kalman smoothing of predicted load cannot be implemented and predictions are not be interpolated between successive load observations.

In the application of the SMLE method to compute mean annual loads for the regional SPARROW models, the number of simulations used to approximate the log likelihood was 1,000 ($S = 1{,}000$). The simulated likelihood function was optimized using the quasi-Newton optimization method as implemented in SAS/IML software (version 12.3, SAS Institute Inc., 2013), with an absolute gradient convergence criterion and a convergence threshold of 1×10^{-6}. If there is no censoring, then the method of estimation defaults to standard maximum likelihood,

with the covariance matrix and bias adjustment derived under the assumption that each individual term in the log-likelihood function (see eq. 2.4) is independent.

Prediction

The prediction methodology makes use of the AR(1) correlation property of the residuals to obtain an estimate of load conditioned on load observations. If a concentration observation is uncensored, then the prediction of daily load for that observation day equals the product of daily flow and the observed concentration, with an adjustment for units of measurement to produce a load estimate in kilograms. If all observations in the sample are uncensored, then conditional prediction is a straightforward application of Kalman-smoothing methods, as described in Lee and others (2016).

As with estimation, the conditioning of predictions on censored observations requires simulation methods. Random realizations of daily loads are generated depending on whether a prediction day contains a censored observation, in which case bounds based on the censoring threshold can be placed on the generating distribution, or no observation is made, implying no bounds are placed on the randomly generated daily load. When an uncensored value of daily load is observed, it then becomes possible to assess the probability of observing that value given the most recent past uncensored observation and the sequence of simulated daily loads generated in the intervening period, thereby establishing a probability for the simulated sequence of intervening daily loads. Multiplying the simulated daily loads by their probability conditioned on the observations and summing for all simulation sequences gives the expected value of daily load. The average annual sum of these expected daily loads over all qualifying years in the prediction period gives the conditional mean annual load, subject to further adjustment to account for skewed uncertainty in the estimate.

In forming conditional probabilities under the assumed AR(1) correlation process, it is necessary to include sample information back in time only as far as the most recent uncensored observation, all information prior to this being superfluous. The specification of the log-likelihood for estimation implies that uncensored observations dictate the points in time when the probability of a simulated sequence of daily loads between these observations can be assessed. Therefore, uncensored observations define statistically independent blocks of time in which daily predictions are determined.

Consider a block of days in which the first day of the block is the first prediction day following a day with an uncensored observation or, in the case of the first block in the prediction period, is the beginning of the prediction period. The last day of the block corresponds to the day of the next uncensored observation in the prediction period or, in the case of the last block in the period, the last day in the prediction period if that day is not also an uncensored observation day. Delineated in this way, the number of blocks B covering a prediction period equals the number of uncensored observations in the prediction period, plus 1 if the last day of the prediction period is not an uncensored observation.

In the following discussion, prediction blocks are indexed by b and prediction days are indexed by t, where the first day of block b is denoted $\underline{T}_b$ and the last day of the block is $\overline{T}_b$. Prediction days may contain gaps due to missing ancillary data, such as daily flow. To account for this, g_t denotes the number of days between prediction day t and prediction day $t-1$. T_b denotes the set of prediction days in block b that are not days containing an uncensored measurement, meaning in most cases T_b excludes the last day, $\overline{T}_b$, and $Y_b = \{y_t : t \in T_b\}$ is the set of log-

transformed concentrations for those days. Lastly, T_b^c denotes the days in block b which contain a censored measurement of concentration, and $C_b = \{\bar{y}_t : y_t \leq \bar{y}_t, t \in T_b^c\}$ represents the set of censored concentrations for those days.

The probability of the sequence of log-transformed concentrations occurring in block b (Y_b) that are unconstrained by uncensored measurement, and thereby exclude the measurement $y_{\overline{T}_b}$, conditioned on the set of available information $\{y_{\overline{T}_b}, C_b, y_{\underline{T}_b-1}\}$ is given by Bayes rule as

$$\Pr\left(Y_b \middle| y_{\overline{T}_b}, C_b, y_{\underline{T}_b-1}\right) = \frac{\Pr\left(Y_b, y_{\overline{T}_b} \middle| C_b, y_{\underline{T}_b-1}\right)}{\Pr\left(y_{\overline{T}_b} \middle| C_b, y_{\underline{T}_b-1}\right)} = \frac{\Pr\left(Y_b, y_{\overline{T}_b} \middle| C_b, y_{\underline{T}_b-1}\right)}{\int \Pr\left(Y_b, y_{\overline{T}_b} \middle| C_b, y_{\underline{T}_b-1}\right) dY_b}, \qquad (2.17)$$

where the integral in the denominator of the last term refers to multivariate integration over the support for Y_b which, in the case of the normal distribution, is the n_b-dimensional interval $\mathbb{R}^{n_b}$, where $\mathbb{R}$ is the real line $(-\infty, \infty)$ and n_b is the number of elements in Y_b. For the conditional probabilities appearing in equation 2.17 and the sequel the conditioning information is implied to include the parameter vector θ.

The numerator of equation 2.17 can be further decomposed as follows:

$$\begin{aligned}
\Pr\left(Y_b, y_{\overline{T}_b} \middle| C_b, y_{\underline{T}_b-1}\right) &= p_{\overline{T}_b}\left(y_{\overline{T}_b} \middle| Y_b, C_b, y_{\underline{T}_b-1}\right) \Pr\left(Y_b \middle| C_b, y_{\underline{T}_b-1}\right) \\
&= p_{\overline{T}_b}\left(y_{\overline{T}_b} \middle| y_{\overline{T}_b-1}\right) \Pr\left(Y_b \middle| C_b, y_{\underline{T}_b-1}\right) \\
&= p_{\overline{T}_b}\left(y_{\overline{T}_b} \middle| y_{\overline{T}_b-1}\right) \frac{\Pr\left(Y_b, C_b \middle| y_{\underline{T}_b-1}\right)}{\Pr\left(C_b \middle| y_{\underline{T}_b-1}\right)} \\
&= p_{\overline{T}_b}\left(y_{\overline{T}_b} \middle| y_{\overline{T}_b-1}\right) \frac{\Pr\left(C_b \middle| Y_b, y_{\underline{T}_b-1}\right) \Pr\left(Y_b \middle| y_{\underline{T}_b-1}\right)}{\Pr\left(C_b \middle| y_{\underline{T}_b-1}\right)}, \qquad (2.18)
\end{aligned}$$

where the first equality follows from an application of Bayes rule, the second equality is a consequence of the AR(1) process whereby the previous prediction day suffices for conditioning on all prior information, and the third and fourth equalities are additional applications of Bayes rule.

The probability $\Pr\left(C_b \middle| Y_b, y_{\underline{T}_b-1}\right)$ appearing in the numerator of the last equality in equation 2.18 has a simple representation. Given the value of concentration, y_t, the event $y_t \leq \bar{y}_t$ is either true or false, implying $\Pr\left(C_b \middle| Y_b, y_{\underline{T}_b-1}\right)$ can be expressed as the product of indicator functions

$$\Pr\left(C_b \middle| Y_b, y_{\underline{T}_b-1}\right) = \prod_{t \in T_b^c} \delta(y_t \leq \bar{y}_t), \qquad (2.19)$$

where $\delta(\cdot)$ is a binary value, being equal to 1 if the argument is true and 0 if it is false. The AR(1) property implies the conditional probability of concentration depends only on the previous value of concentration, all other prior information being superfluous. Consequently, the term $\Pr\left(Y_b \middle| y_{\underline{T}_b-1}\right)$ in the numerator of the last equality in equation 2.18 can be expressed as the product of simple conditional marginal probabilities,

$$\Pr\left(Y_b \middle| y_{\underline{T}_b-1}\right) = \prod_{t \in T_b} p_t(y_t | y_{t-1}). \qquad (2.20)$$

Given the relations in equations 2.18 through 2.20, the probability expressed in equation 2.17 of the sequence of daily concentrations in a block conditioned on the measurements $\{y_{\underline{T}_b}, C_b, y_{\underline{T}_b-1}\}$ has the equivalent expression

$$\Pr\left(Y_b \middle| y_{\overline{T}_b}, C_b, y_{\underline{T}_b-1}\right) = \frac{\prod_{t=\underline{T}_b}^{\overline{T}_b} p_t(y_t|y_{t-1}) \prod_{t\in T_b^c} \delta(y_t \le \bar{y}_t)}{\int \prod_{t=\underline{T}_b}^{\overline{T}_b} p_t(y_t|y_{t-1}) \prod_{t\in T_b^c} \delta(y_t \le \bar{y}_t)\, dY_b}. \tag{2.21}$$

Note that the term $\Pr\left(C_b \middle| y_{\underline{T}_b-1}\right)$ in the denominator of the last equality in equation 2.18 does not depend explicitly on Y_b and therefore cancels from the numerator and denominator of equation 2.21.

If the last block in the prediction period does not terminate with a measured concentration, the probability for the block concentrations is

$$\Pr\left(Y_B \middle| C_B, y_{\underline{T}_B-1}\right) = \frac{\Pr\left(C_B \middle| Y_b, y_{\underline{T}_B-1}\right)\Pr\left(Y_b \middle| y_{\underline{T}_B-1}\right)}{\Pr\left(C_B \middle| y_{\underline{T}_B-1}\right)}$$

$$= \frac{\prod_{t=\underline{T}_B}^{\overline{T}_B} p_t(y_t|y_{t-1}) \prod_{t\in T_B^c} \delta(y_t \le \bar{y}_t)}{\int \prod_{t=\underline{T}_B}^{\overline{T}_B} p_t(y_t|y_{t-1}) \prod_{t\in T_B^c} \delta(y_t \le \bar{y}_t)\, dY_B}, \tag{2.22}$$

which matches the formulation shown in equation 2.21. Thus, equation 2.21 is valid for all blocks comprising the prediction period.

The expectation of untransformed concentration y_k^* for a given prediction day k in block b is obtained by multiplying log-transformed concentration by the joint probability in equation 2.21 and integrating over the elements of Y_b,

$$E\left[y_k^* \middle| y_{\overline{T}_b}, C_b, y_{\underline{T}_b-1}\right] = \frac{\int \exp(y_k) \prod_{t=\underline{T}_b}^{\overline{T}_b} p_t(y_t|y_{t-1}) \prod_{t\in T_b^c} \delta(y_t \le \bar{y}_t)\, dY_b}{\int \prod_{t=\underline{T}_b}^{\overline{T}_b} p_t(y_t|y_{t-1}) \prod_{t\in T_b^c} \delta(y_t \le \bar{y}_t)\, dY_b}. \tag{2.23}$$

The same equation applies if $y_{\overline{T}_b}$ is not measured in the last block, where $b = B$. If k refers to a day with an uncensored measurement, then y_k is set equal to the measurement, $y_{\overline{T}_b}$.

Because a block may contain multiple censored measurements, the integration step used for taking the expectation and for normalizing the probability (see the numerator and denominator of eq. 2.23) becomes numerically burdensome and necessitates the use of simulation methods, as are used in the evaluation of the likelihood function for model estimation. In generating simulated outcomes for concentration $\tilde{y}_t^{(s)}$, two generating distributions are used. For prediction days with no measurement, concentrations are generated from the conditional normal probability density, and for days with a censored measurement a specific value of concentration is drawn from a conditional truncated normal distribution,

$$\tilde{y}_t^{(s)} \sim \begin{cases} p_t\left(y \middle| \tilde{y}_{t-1}^{(s)}\right), \text{ if } t \in T_b \text{ and } t \notin T_b^c, \\ \\ \delta(y \le \bar{y}_t)\, p_t\left(y \middle| \tilde{y}_{t-1}^{(s)}\right) / P_t\left(\bar{y}_t \middle| \tilde{y}_{t-1}^{(s)}\right), \text{ if } t \in T_b^c, \end{cases} \tag{2.24}$$

where the delta function ($\delta(\cdot)$) is the distribution for censored observations, which specifies that the probability is 1 inside the censoring range for y and 0 outside the range. In application, the simulated values are generated using the standard normal probability density and the standard uniform density, according to

$$\tilde{y}_t^{(s)} = \tilde{z}_t^{(s)} + \hat{\sigma}_t \begin{cases} e_t^{(s)} \sim N(0,1), \text{ if } t \in T_b \text{ and } t \notin T_b^c, \\ \Phi^{-1}\left(u_t^{(s)} \Phi\left(\frac{\bar{y}_t - \tilde{z}_t^{(s)}}{\hat{\sigma}_t} \right) \right), \; u_t^{(s)} \sim U(0,1) \text{ if } t \in T_b^c, \end{cases} \tag{2.25}$$

where $\tilde{z}_t^{(s)} = X_t\hat{\beta} + \hat{\rho}_t\left(\tilde{y}_{t-1}^{(s)} - X_{t-1}\hat{\beta}\right)$, $\hat{\rho}_t = \hat{\rho}^{g_t/d_{\text{Median}}}$, and $\hat{\sigma}_t = \hat{\sigma}\sqrt{1-\hat{\rho}_t^2}$, and $\hat{\beta}$, $\hat{\rho}$, and $\hat{\sigma}$ are the simulated maximum likelihood estimates of the model parameters, including the first-order bias adjustment. For the first prediction day in a block with a prior uncensored measurement, the simulated conditioning value of log concentration is given by the measured value, so that $\tilde{y}_{\underline{T}_b-1}^{(s)} = y_{\underline{T}_b-1}$, for all $s = 1, \ldots, S$. For the first prediction day of the first block ($b = 1$), there is no prior information and $\tilde{y}_{\underline{T}_1-1}^{(s)} = 0$. Additionally, for this day $\hat{\rho}_t = 0$, implying that $\hat{\sigma}_t = \hat{\sigma}$. The delta function appearing in equation 2.24 used to simulate censored observations is not required in equation 2.25. As pointed out by Lee (1999), an advantage of using the uniform random variable to simulate log concentration on days containing a censored observation is that all simulated values are admissible, meaning that each simulated value adheres to the censoring condition $y_t \leq \bar{y}_t$.

The estimate of the probability of the simulation sequence, analogous to the conditional probability shown in equation 2.21, is given by

$$\widehat{\Pr}\left(\tilde{Y}_b^{(s)} \middle| y_{\overline{T}_b}, C_b, y_{\underline{T}_b-1}\right) =$$

$$\begin{cases} \dfrac{p_{\overline{T}_b}\left(y_{\overline{T}_b} \middle| \tilde{y}_{\overline{T}_b-1}^{(s)}\right) \prod_{t \in T_b^c} P_t\left(\bar{y}_t \middle| \tilde{y}_{t-1}^{(s)}\right)}{S^{-1}\sum_{s=1}^{S} p_{\overline{T}_b}\left(y_{\overline{T}_b} \middle| \tilde{y}_{\overline{T}_b-1}^{(s)}\right) \prod_{t \in T_b^c} P_t\left(\bar{y}_t \middle| \tilde{y}_{t-1}^{(s)}\right)}, \text{ if } y_{\overline{T}_b} \text{observed}, \\ \dfrac{\prod_{t \in T_b^c} P_t\left(\bar{y}_t \middle| \tilde{y}_{t-1}^{(s)}\right)}{S^{-1}\sum_{s=1}^{S} \prod_{t \in T_b^c} P_t\left(\bar{y}_t \middle| \tilde{y}_{t-1}^{(s)}\right)}, \text{ if } y_{\overline{T}_b} \text{not observed } (b = B), \end{cases} \tag{2.26}$$

where $s = 1, \ldots, S$, $\tilde{Y}_b^{(s)} = \left\{\tilde{y}_t^{(s)} : t \in T_b\right\}$ is the sequence of simulated log-transformed concentrations in block b, $p_t\left(y_{\overline{T}_b} \middle| \tilde{y}_{\overline{T}_b-1}^{(s)}\right) = \hat{\sigma}_{\overline{T}_b}^{-1}\phi\left(\left(y_{\overline{T}_b} - \tilde{z}_{\overline{T}_b}^{(s)}\right)/\hat{\sigma}_{\overline{T}_b}\right)$ is a conditional normal probability density, and $P_t\left(\bar{y}_t \middle| \tilde{y}_{t-1}^{(s)}\right) = \Phi\left(\left(y_{\overline{T}_b} - \tilde{z}_{\overline{T}_b}^{(s)}\right)/\hat{\sigma}_{\overline{T}_b}\right)$ is the cumulative normal probability. Technically, to be a probability the sum across all s should be 1, whereas the sum across all values of s of equation 2.26 is S. Consequently, the probability described in equation 2.26 is referred to as a scaled probability. The scaling by S is used to induce numerical stability among the individual probability terms (Lee, 1999) and is removed in the final step of simulating the expectation of daily values.

Given the simulated scaled probability in equation 2.26, the simulated conditional expectation of untransformed concentration y_k^* for day k, for some block b is given by

$$\hat{E}\left[y_k^* \middle| y_{\overline{T}_b}, C_b, y_{\underline{T}_b-1}\right] = \frac{S^{-1}\sum_{s=1}^{S}\exp\left(\tilde{y}_k^{(s)}\right)p_{\overline{T}_b}\left(y_{\overline{T}_b}\middle|\tilde{y}_{\overline{T}_b-1}^{(s)}\right)\prod_{t\in T_b^c}P_t\left(\bar{y}_t\middle|\tilde{y}_{t-1}^{(s)}\right)}{S^{-1}\sum_{s=1}^{S}p_{\overline{T}_b}\left(y_{\overline{T}_b}\middle|\tilde{y}_{\overline{T}_b-1}^{(s)}\right)\prod_{t\in T_b^c}P_t\left(\bar{y}_t\middle|\tilde{y}_{t-1}^{(s)}\right)}, \quad (2.27)$$

where, rather than taking a simple sum in the numerator, the normalization by S has been removed by taking the average over s of the product of simulated concentration and the simulation probability (the inclusion of the S^{-1} term in the numerator of eq. 2.27). If b refers to the last block in the prediction period ($b = B$) and the last prediction day of the last block is not measured or is measured with censoring, then the conditional marginal probability $p_{\overline{T}_b}\left(y_{\overline{T}_b}\middle|\tilde{y}_{\overline{T}_b-1}^{(s)}\right)$ is dropped from both the numerator and the denominator.

Under general conditions, the limits as S approaches infinity of the averages representing the numerator and denominator of equation 2.27 go to their respective expectations (Amemiya, 1985). Therefore, the approximation in equation 2.27 can be evaluated by taking the expectation of the numerator and denominator separately, recognizing the mechanisms by which the simulated values $\tilde{y}_{t-1}^{(s)}$ are generated (eq. 2.24), and comparing the ratio of these expectations to the theoretical conditional expectation given in equation 2.23. The expectation of the numerator and denominator terms of equation 2.27 are given by

$$E\left[\exp\left(\tilde{y}_k^{(s)}\right)\prod_{t\in T_b^c}P_t\left(\bar{y}_t\middle|\tilde{y}_{t-1}^{(s)}\right)\right] =$$

$$\int\left(\exp(y_k)\prod_{t\in T_b^c}P_t(\bar{y}_t|y_{t-1})\frac{\delta(y_t\le\bar{y}_t)p_t(y_t|y_{t-1})}{P_t(\bar{y}_t|y_{t-1})}\right)\left(\prod_{t\in T_b\setminus T_b^c}p_t(y_t|y_{t-1})\right)dY_b$$

$$= \int\exp(y_k)\prod_{t\in T_b^c}\delta(y_t\le\bar{y}_t)\prod_{t\in T_b}p_t(y_t|y_{t-1})\,dY_b, \quad (2.28)$$

and

$$E\left[\prod_{t\in T_b^c}P_t\left(\bar{y}_t\middle|\tilde{y}_{t-1}^{(s)}\right)\right] =$$

$$\int\left(\prod_{t\in T_b^c}P_t(\bar{y}_t|y_{t-1})\frac{\delta(y_t\le\bar{y}_t)p_t(y_t|y_{t-1})}{P_t(\bar{y}_t|y_{t-1})}\right)\left(\prod_{t\in T_b\setminus T_b^c}p_t(y_t|y_{t-1})\right)dY_b$$

$$= \int\prod_{t\in T_b^c}\delta(y_t\le\bar{y}_t)\prod_{t\in T_b}p_t(y_t|y_{t-1})\,dY_b, \quad (2.29)$$

which exactly match the numerator and denominator of equation 2.23. Therefore, the simulation evaluation function for block b given in equation 2.27 combined with the simulation generating functions for censored observation days and non-measurement days (eq. 2.25) represent a consistent estimate of the expectation of untransformed concentration conditioned on censored and uncensored sample measurements.

As with the simulation of weights used to evaluate the likelihood function in model estimation, the scaled probabilities in equation 2.26 can be generated recursively. The initial scaled probabilities for each block b and simulation s, denoted $\widehat{\Pr}_{0b}^{(s)}$, are set equal to 1. Cycling forward in time through each prediction day in block b, simulated values of $\tilde{y}_t^{(s)}$ are generated according to equation 2.25. If a measurement day is encountered, the updated scaled probability, denoted $\widehat{\Pr}_b^{(s)}$, which becomes the new initial value $\widehat{\Pr}_{0b}^{(s)}$ for the next encountered measurement day, is given by

$$\widehat{\Pr}_b^{(s)} = \frac{\Pr_t\left(\omega_t \middle| \tilde{y}_{t-1}^{(s)}\right)\widehat{\Pr}_{0b}^{(s)}}{S^{-1}\sum_{s=1}^{S}\Pr_t\left(\omega_t \middle| \tilde{y}_{t-1}^{(s)}\right)\widehat{\Pr}_{0b}^{(s)}}, \tag{2.30}$$

where ω_t is the measurement on date t, and $\Pr_t\left(\omega_t \middle| \tilde{y}_{t-1}^{(s)}\right)$ is the conditional probability of that measurement given the previous prediction day's simulated log concentration for simulation sequence s. If $\omega_t = \bar{y}_t$, implying the measurement is a censored observation, then $\Pr_t\left(\omega_t \middle| \tilde{y}_{t-1}^{(s)}\right)$ is set equal to the conditional normal cumulative probability $P_t\left(\bar{y}_t \middle| \tilde{y}_{t-1}^{(s)}\right)$; conversely, if $\omega_t = y_t$, so that the measurement is uncensored and the observation pertains to the last observation in the block, then $\Pr_t\left(\omega_t \middle| \tilde{y}_{t-1}^{(s)}\right)$ equals the conditional normal density function $p_t\left(y_t \middle| \tilde{y}_{t-1}^{(s)}\right)$.

Detrended Loads for SPARROW Application

SPARROW modeling requires load estimates to be detrended to a common base year (Schwarz and others, 2006). There are two sources of trend in the load estimation model: trend in flow, and direct trend affecting water-quality concentrations for a given flow. Trend in flow affects trend in load in two ways: through the effect of flow on concentration, and through the multiplication of concentration by flow to determine load. The direct trend in concentration is accounted for in the estimation of the water-quality model (eq. 2.1) by including a linear time trend explanatory variable. In prediction, to compute a detrended load estimate, the time trend variable is set equal to the normalization date (June 30, 2012) for the entire prediction period. To correct for trend in flow, it is necessary to separately estimate a detrended flow variable, this variable then being substituted for the flow variables determining concentration and load.

The detrended flow variable is obtained by regressing the logarithm of daily flow q_t on a constant, a decimal time trend variable D_t, and the sine and cosine of the product of 2π and decimal time to account for seasonal variations in flow. The residuals of the model are assumed to adhere to a 30-lag autoregressive (AR(30)) correlation structure, which accounts for significant lagged correlation. Given estimates of the linear trend coefficient, detrended log-transformed flow $\bar{q}_t$ for day t is given by

$$\bar{q}_t = q_t + \alpha_D(D_0 - D_t), \tag{2.31}$$

where α_D is the coefficient on decimal time in the flow equation, and D_0 is the value of decimal time for the normalization date.

Detrended log-transformed load for day k is computed by replacing the $\tilde{y}_k^{(s)}$ term in equation 2.27 with an adjusted value $\tilde{\tilde{y}}_k^{(s)}$ given by

$$\tilde{\tilde{y}}_k^{(s)} = \tilde{y}_k^{(s)} + \ln \kappa + \bar{q}_k + \beta_q(\bar{q}_k - q_t) + \beta_D(D_0 - D_k), \tag{2.32}$$

where κ is a constant that converts the units of concentration-flow to units of load, and β_q and β_D are the coefficients for logarithm of flow and decimal time in the water-quality model (eq. 2.1). Note that the detrended flow ($\bar{q}_k$) without a coefficient appearing in equation 2.32 represents the flow that gets multiplied by predicted concentration to determine detrended load for day k.

For the purposes of this report, the prediction period was determined to be too short for flow to have a quantitatively significant effect on trends in load. Consequently, in deriving detrended load the detrended logarithm of flow was set to equal the logarithm of actual flow. Detrending with respect to decimal time in the water-quality model is done only if certain conditions are met in addition to those that qualify a station for inclusion in the SPARROW model. The additional conditions require the water-quality record to be no shorter than 3 years and end no more than 2.25 years before the normalization date. If these conditions are not met, then the decimal time variable is generally dropped from the water-quality model. The only exception is if the water-quality record extends more than 50 percent of the flow record; however, in this case although decimal time is retained in the water-quality model the decimal-time term is excluded from the adjustment in equation 2.32 so that detrending directly with respect to time in the water-quality model is not implemented.

Aggregation and Correction for Bias

The detrended mean annual load estimate is obtained by aggregating the detrended daily load estimates described above. Not all daily load predictions over the prediction period are included in the aggregation. Only prediction days that are in a water year in which all days of the year have flow data are used in aggregation. This ensures that all days in a water year are equally represented in the mean annual load estimate. A water year is the 12-month period October 1 through September 30 designated by the calendar year in which it ends.

Although the mean annual load estimate is restricted to years with only complete flow records, predictions for days contained in water years with only partial flow data are not excluded from the prediction process. The prediction algorithm described in equation 2.27 (with detrending through application of eq. 2.32) includes all days in the prediction period with valid flow data. Gaps in the flow record are accounted for by reducing the serial correlation between successive observations in the flow record according to the same approach used to account for gaps between successive water-quality observations in estimating the water-quality model. This means that water-quality observations that occur in water years with an incomplete flow record are appropriately used to condition the load prediction, even though those observations are not directly included in the mean annual load estimate. As a last step, the mean annual load is obtained by normalizing the aggregated prediction days by the number of water years with complete flow records, that is, by the number of years covered by the predictions included in the aggregation.

If $\bar{T}_b$ represents the set of prediction days in block b that meet the criterion for inclusion in mean load, detrended mean annual load $\bar{L}^*$ is given by

$$\bar{L}^* = \bar{N}^{-1} S^{-1} \sum_{s=1}^{S} \sum_{b=1}^{B} \hat{\mathrm{P}}_b^{(s)} \sum_{k \in \bar{T}_b} \exp\left(\tilde{\tilde{y}}_k^{(s)}\right), \tag{2.33}$$

where $\bar{N}$ is the number of years in the prediction period with complete flow data, and $\hat{\mathrm{P}}_b^{(s)}$ is the scaled probability, $\widehat{\mathrm{Pr}}_b^{(s)}$ (see eq. 2.26), after iterating over all water-quality observations in block b.

The load estimation model is a nonlinear function of the model coefficients, θ. Models that are linear in the coefficients are unbiased (Amemiya, 1985), implying predictions are unbiased. The SMLE coefficient estimates $\hat{\theta}$ of the nonlinear model are consistent and in large samples have a symmetric, normal distribution; however, due to curvature effects induced by the model functional form, the statistical variation of the randomly estimated coefficients do not necessarily balance between low and high random values of the estimates, causing prediction bias to have potentially greater sensitivity to sampling uncertainty in the coefficient estimates.

To account for this bias, a method of bias correction is used based on a parametric bootstrap (Schwarz and others, 2006). Multiple values of the model coefficients $\hat{\theta}^{(r)}$, $r = 1, \ldots, R$, are randomly generated using a multivariate normal distribution having a mean equal to the adjusted SMLE coefficient estimates and a covariance given by the SMLE asymptotic covariance matrix. The randomly generated coefficients are constrained according to the same constraints applied in model estimation ($0.999 \geq \hat{\rho} \geq 0$, and $\hat{\sigma} \geq 0.01$).

The prediction algorithm described above is implemented for each set of coefficients, with daily predictions appropriately aggregated to obtain a mean annual load estimate, thereby giving R estimates of mean annual load. The same random values used to represent the model residuals are used for all bootstrap iterations, so the only source of variation across the R predictions is due to random variations in the coefficients. For purposes of this report, the number of randomly generated sets of coefficients used in the bootstrap (R) is the same as number of simulations (S) used to evaluate the integrals described in equation 2.27.

The bias adjustment applied here corrects relative bias, which is bias expressed as the ratio of the estimate to the actual value. This approach has the advantage that bias-corrected load estimates cannot become negative, unlike with an additive bias correction. A disadvantage, however, is that daily loads also corrected for relative bias, if averaged, do not equal the bias-corrected mean annual load. This is a consequence of the nonlinear form of a relative bias; an additive bias correction would not affect the equality. Note that it is not necessary to correct for retransformation bias, as is done in other methods such as that described by Cohn (2005). The prediction of mean annual load ($\bar{L}^*$) produced from equation 2.27 represents an integration of randomly realized load expressed in real space, appropriately weighted by the probability of the load outcome. Thus, retransformation bias is fully accounted for in the prediction method, and no additional adjustment is required.

According to the bootstrap paradigm, the relative bias ratio can be estimated as the ratio of the average of the R bootstrap estimates of mean annual load to the mean annual load estimated from the adjusted-SMLE coefficient estimates. Thus, if $\bar{L}^*(\theta)$ is defined to be the untransformed mean detrended load expressed as a function of an arbitrary coefficient vector θ, then the bias corrected estimated mean detrended load is

$$\hat{\bar{L}}^* = \frac{\bar{L}^*\left(\hat{\theta}\right)^2}{R^{-1} \sum_{r=1}^{R} \bar{L}^*\left(\hat{\theta}^{(r)}\right)}. \tag{2.34}$$

Variance Estimation

There are two sources of error in a model prediction: model error, which is the error arising from the model residual ε_t, and sampling error, which is the random error induced by basing estimates of model coefficients on a randomly generated finite sample. Because the model coefficients are consistent, the sampling error approaches 0 as sample size gets large. Conversely, model error is insensitive to sample size, although its importance in prediction uncertainty is diminished the greater the temporal aggregation of a prediction; for example, model error is more important for error in the prediction of daily load than in the prediction of mean annual load. In assessing prediction error for detrended load, a simplifying assumption is made that detrended flow has no uncertainty. This is not technically true because there is sampling error in the estimated value of the α_D coefficient. However, there are two reasons to support ignoring this error: the flow model is based on a much larger set of observations than the water-quality model, and the trend coefficient in the flow model is superefficient, in the sense that the standard error of this coefficient approaches 0 according to the inverse of the length of the number of observations in the flow record as opposed to the square root of the number of observations, as is generally the case with non-trending variables (Schwarz and others, 2006).

The model error component of prediction error is computed for coefficients set at the adjusted SMLE values. Because conditioning causes predictions between uncensored observations of different blocks to be independent, the variance of the aggregated daily predictions is the sum across blocks of the variance of the aggregation within each block, weighted by the scaled probability. Thus, the model error component of variance $V_m(\bar{L}^*)$ is given by

$$V_m(\bar{L}^*) = \bar{N}^{-2} S^{-1} \sum_{s=1}^{S} \sum_{b=1}^{B} \hat{\mathrm{P}}_b^{(s)} \left(\sum_{k \in \bar{T}_b} \exp\left(\tilde{\tilde{y}}_k^{(s)}\right) - \bar{L}_b^* \right)^2, \tag{2.35}$$

where $\bar{L}_b^* = S^{-1} \sum_{s=1}^{S} \hat{\mathrm{P}}_b^{(s)} \sum_{k \in \bar{T}_b} \exp\left(\tilde{\tilde{y}}_k^{(s)}\right)$ is the expected detrended load for valid prediction days contained in block b.

The sampling error component of prediction error is assessed using a parametric bootstrap method, based on the same bootstrap outcomes generated to correct for nonlinear bias, as described in the "Aggregation and Correction for Bias" section of this appendix. The sampling error component variance $V_s(\bar{L}^*)$ is just the variance of the bootstrap outcomes of mean detrended load,

$$V_s(\bar{L}^*) = R^{-1} \sum_{r=1}^{R} \left(\bar{L}^*\left(\hat{\theta}^{(r)}\right) - \overline{\bar{L}^*\left(\hat{\theta}^{(r)}\right)} \right)^2, \tag{2.36}$$

where $\overline{\bar{L}^*\left(\hat{\theta}^{(r)}\right)} = R^{-1} \sum_{r=1}^{R} \bar{L}^*\left(\hat{\theta}^{(r)}\right)$ is the average of the bootstrap load estimates.

Combining the two components of variance, the variance of the mean annual detrended load estimate $\mathrm{V}(\bar{L}^*)$ is given by

$$\mathrm{V}(\bar{L}^*) = V_s(\bar{L}^*) + V_m(\bar{L}^*). \tag{2.37}$$

The algorithm used to form a Kalman-smoothed estimate of mean annual load and its variance arising from model error, conditioned on known values of the coefficients, is detailed in the "Algorithm for the Conditional Prediction of Detrended Mean Annual Load" section of this appendix.

Algorithm for the Evaluation of the Log-Likelihood Function Using the Simulation Method

The log-likelihood estimation algorithm uses the following variable definitions:

N is the number of observations;

i is the observation index, $i = 1, \ldots, N$;

S is the number of simulation iterations;

s is the simulation iteration index, $s = 1, \ldots, S$;

d_i is the number of days between observations i and $i - 1$;

d_{Med} is the median number of days between observations;

y_i is the uncensored value of the logarithm of contaminant concentration;

$\bar{y}_i$ is the logarithm of the contaminant concentration censoring threshold;

X_i is the $1 \times K$ row vector of the water-quality model explanatory variables (including an intercept);

β is the $K \times 1$ vector of coefficients associated with the explanatory variables;

ρ is the daily serial correlation parameter, to be estimated and expressed as the correlation over the median number of days between observations in the sample (d_{Med});

σ_ε is the standard deviation of the random innovation (see eq. 2.2);

σ is the standard deviation of the unconditional residual and equal to $\sigma_\varepsilon / \sqrt{1 - \rho^2}$;

$\boldsymbol{r_i}$ is the serial correlation between observations $\boldsymbol{i}$ and $\boldsymbol{i - 1}$;

$\boldsymbol{\sigma_i}$ is the standard deviation of the concentration model residual conditioned on concentration in period $\boldsymbol{i - 1}$ (unconditioned for the first observation);

$u_i^{(s)}$ is a random variable for observation i and simulation iteration s, drawn from a uniform distribution over the range 0 to 1 ($U(0,1)$);

$\tilde{y}_i^{(s)}$ is the randomly generated value of the logarithm of concentration for observation i and simulation iteration s, simulated from a truncated normal conditional distribution for censored observations but initialized at the fixed value 0 and reinitialized for uncensored observations by y_i (see eq. 2.13);

$\hat{w}_i^{(s)}$ is the simulated value of the probability weight for observation i and simulation iteration s (see eq. 2.12);

$\widehat{\Pr}_i$ is the simulation estimate of the probability of the i-th observation, either censored or uncensored, conditioned on all prior observations (see eq. 2.6); and

L is the log-likelihood.

The algorithm used to evaluate the log-likelihood function for a given set of model coefficients $\{\boldsymbol{\beta}, \boldsymbol{\rho}, \boldsymbol{\sigma_\varepsilon}\}$ is given by:

Initialize:

$\hat{w}_0^{(s)} = 1$ and $\tilde{y}_0^{(s)} = 0$, $s = 1, \ldots, S$;

$u_i^{(s)} \sim U(0,1)$, $s = 1, \ldots, S$ and $i = 1, \ldots, N$;

$L = 0$, $X_0 = 0$ and $r_1 = 0$;

$r_i = \rho^{d_i / d_{\text{Med}}}$, $i = 2, \ldots, N$;

$\sigma_i = \sigma \sqrt{1 - r_i^2}$, $i = 1, \ldots, N$;

Do $i = 1$ to N:

$z_i^{(s)} = (X_i - r_i X_{i-1})\beta + r_i \tilde{y}_{i-1}^{(s)}$, $s = 1, \ldots S$;

If observation i is censored:

$$f^{(s)} = \Phi\left(\left(\bar{y}_i - z_i^{(s)}\right)/\sigma_i\right), s = 1, \ldots S;$$
$$\tilde{y}_i^{(s)} = z_i^{(s)} + \sigma_i \Phi^{-1}\left(u_i^{(s)} f^{(s)}\right), s = 1, \ldots S;$$
$$\widehat{\Pr}_i = S^{-1} \sum_{s=1}^{S} \hat{w}_{i-1}^{(s)} f^{(s)};$$
$$\hat{w}_i^{(s)} = \hat{w}_{i-1}^{(s)} f^{(s)} / \widehat{\Pr}_i, s = 1, \ldots S;$$

Else:

$$f^{(s)} = \phi\left(\left(y_i - z_i^{(s)}\right)/\sigma_i\right)/\sigma_i, s = 1, \ldots S;$$
$$\widehat{Pr}_i = S^{-1} \sum_{s=1}^{S} \hat{w}_{i-1}^{(s)} f^{(s)};$$
$$\tilde{y}_i^{(s)} = y_i, s = 1, \ldots S;$$
$$\hat{w}_i^{(s)} = 1, s = 1, \ldots S;$$
$$L = L + \ln \widehat{\Pr}_i;$$

Algorithm for the Conditional Prediction of Detrended Mean Annual Load

The simulation method algorithm used to estimate the conditional prediction of detrended mean annual load uses the following variable definitions:

T is the number of days in the prediction period with valid explanatory variable data;

t indexes successive days with valid explanatory variable data in the prediction period, $t = 1, \ldots, T$;

$\bar{N}$ is the number of valid water years containing a complete year of explanatory variable observations;

B is the number of blocks of days between successive uncensored observations, plus 1 if the last day of the prediction period is not an uncensored observation;

b indexes blocks of days between successive uncensored observations, with $b = 1, \ldots, B$;

T_b is the number of days with valid explanatory variable data in block b;

i is the successive day of flow in a block, $i = 1, \ldots, T_b$;

$\bar{T}_b$ is the set of prediction days in block b that are part of a water year with a complete flow record;

S is the number of simulation iterations;

s indexes simulations, with $s = 1, \ldots, S$;

y_t is the uncensored value of the logarithm of contaminant concentration for day t, if known;

$\bar{y}_t$ is the logarithm of the contaminant concentration censoring threshold for day t if day t is an observation day containing a censored observation;

X_t is the row vector of explanatory variables for day t (including an intercept);

β is a vector of coefficients associated with the explanatory variables;

q_t is the logarithm of daily flow for day t;

$\bar{q}_t$ is the detrended logarithm of daily flow for day t;

D_t is decimal time for day t;

D_0 is the decimal normalization date;

β_q is the water-quality model coefficient associated with the logarithm of daily flow;

β_D is the water-quality model coefficient associated with the decimal time;

d_{Median} is the median number of days between observations in the estimation sample;

d_t is the number of days between successive days t and $t-1$ with valid explanatory variable data;

ρ is the serial correlation parameter expressed as the correlation over the median number of days between observations in the sample d_{Median};

σ is $\sigma_\varepsilon/\sqrt{1-\rho^2}$ the standard deviation of the unconditional residual, where σ_ε is the standard deviation of a daily innovation to the water-quality model residual;

r_t is the serial correlation between observations t and $t-1$;

σ_t is the standard deviation of the concentration model residual conditioned on concentration in period $t-1$ (unconditioned for the first observation);

$u_t^{(s)}$ is a random variable for day t and simulation s, drawn from a uniform distribution ($U(0,1)$) if day t corresponds to a prediction day with a censored concentration measurement, otherwise $u_t^{(s)}$ is drawn from a standard normal distribution;

$\tilde{y}^{(s)}$ is the simulated daily logarithm of concentration for simulation s;

$\widehat{\Pr}_b^{(s)}$ is the scaled probability for the simulation s outcomes in block b, conditioned on concentration measurements in the block (see eq. 2.30);

$L_b^{(s)}$ is simulation s accumulated load for block b;

$\bar{L}_b^*$ is accumulated detrended load for valid water-year prediction days in block b;

$\bar{L}^*$ is detrended mean annual load for valid water years (see eq. 2.33); and

$V_m(\bar{L}^*)$ is the variance of the model component of prediction error for $\bar{L}^*$ (see eq. 2.35).

The algorithm for the prediction of mean annual load conditioned on infrequent censored or uncensored observations of concentration within the prediction period, and on a given set of values of the model coefficients $\{\beta, \rho, \sigma_\varepsilon\}$ is given by:

Initialize:

$\widehat{\Pr}_b^{(s)} = 1$ and $L_b^{(s)} = 0$, $s = 1, \ldots, S$ and $b = 1, \ldots, B$;

$u_t^{(s)} \sim U(0,1)$, $s = 1, \ldots, S$, for all t containing a censored concentration, and $u_t^{(s)} \sim N(0,1)$, $s = 1, \ldots, S$, for all other t;

$\tilde{y}^{(s)} = 0$, $s = 1, \ldots, S$;

$r_1 = 0$ and $r_t = \rho^{d_t/d_{\text{Median}}}$, $t = 2, \ldots, T$;

$\sigma_t = \sigma\sqrt{1-r_t^2}$, $t = 1, \ldots, T$;

$T_0 = 0$.

Do $b = 1$ to B:

Do $t = T_0 + 1$ to $\bar{T}_b$:

$z^{(s)} = (X_t - r_t X_{t-1})\beta + r_t \tilde{y}^{(s)}$, $s = 1, \ldots S$;

If t has no concentration measurement:

$\tilde{y}^{(s)} = z^{(s)} + \sigma_t u_t^{(s)}$, $s = 1, \ldots, S$;

Else:

If t has a censored concentration measurement:

$f^{(s)} = \Phi\left(\left(\bar{y}_t - z^{(s)}\right)/\sigma_t\right)$, $s = 1, \ldots S$;

$\tilde{y}^{(s)} = z^{(s)} + \sigma_t \Phi^{-1}\left(u_t^{(s)} f^{(s)}\right)$, $s = 1, \ldots S$;

Else:

$f^{(s)} = \phi\left(\left(\bar{y}_t - z^{(s)}\right)/\sigma_t\right)/\sigma_t$, $s = 1, \ldots S$;

$\tilde{y}^{(s)} = y_t$, $s = 1, \ldots S$;

$$\widehat{\Pr}_b^{(s)} = \widehat{\Pr}_b^{(s)} f^{(s)}, s = 1, \ldots S;$$
$$\widehat{\Pr}_b^{(s)} = \widehat{\Pr}_b^{(s)} \Big/ \left(S^{-1} \sum_{s=1}^{S} \widehat{\Pr}_b^{(s)} \right), s = 1, \ldots S;$$

If $t \in \bar{T}_b$ then aggregate detrended daily load:

$$L_b^{(s)} = L_b^{(s)} + \exp\left(\tilde{y}^{(s)} + \bar{q}_t + \beta_q (\bar{q}_t - q_t) + \beta_D (D_0 - D_t) \right), s = 1, \ldots S;$$
$$T_0 = T_0 + \bar{T}_b;$$

Complete the estimation of detrended mean annual load and its model error:

$$\bar{L}_b^* = S^{-1} \sum_{s=1}^{S} \widehat{\Pr}_b^{(s)} L_b^{(s)};$$
$$\bar{L}^* = \bar{N}^{-1} \sum_{b=1}^{B} \bar{L}_b^*;$$
$$V_m(\bar{L}^*) = \bar{N}^{-2} S^{-1} \sum_{s=1}^{S} \sum_{b=1}^{B} \widehat{\Pr}_b^{(s)} \left(L_b^{(s)} - \bar{L}_b^* \right)^2;$$

References Cited

Amemiya, T., 1973, Regression analysis when the dependent variable is truncated normal: Econometrica, v. 41, no. 6, p. 997–1016. [Also available at https://doi.org/10.2307/1914031.]

Amemiya, T., 1985, Advanced econometrics: Cambridge, Mass., Harvard University Press, 536 p.

Andrews, D.W.K., 1992, Generic uniform convergence: Econometric Theory, v. 8, no. 2, p. 241–257. [Also available at https://doi.org/10.1017/S0266466600012780.]

Andrews, D.W.K., and McDermott, C.J., 1995, Nonlinear models with deterministically trending variables: The Review of Economic Studies, v. 62, no. 3, p. 343-360. [Also available at https://doi.org/10.2307/2298032.]

Cohn, T.A., 2005, Estimating contaminant loads in rivers—An application of adjusted maximum likelihood to type 1 censored data: Water Resources Research, v. 41, no. 7, p. 1–13. [Also available at https://doi.org/10.1029/2004WR003833.]

Conte, S.D. and de Boor, C, 1972, Elementary numerical analysis—An algorithmic approach (2d ed.): New York, McGraw-Hill, 396 p.

Davidson, J., 1994, Stochastic limit theory: Oxford, Oxford University Press, 539 p.

Davidson, R., and MacKinnon, J.G., 1993, Estimation and inference in econometrics: Oxford, Oxford University Press, 875 p.

Harvey, A.C., 1989, Forecasting, structural time series models and the Kalman filter: Cambridge, U.K., Cambridge University Press, 554 p.

Gallant, A.R., and White, H., 1988, A unified theory of estimation and inference for nonlinear dynamic models: Oxford, Basil Blackwell, 155 p.

Lee, L.-F., 1999, Estimation of dynamic and ARCH Tobit models: Journal of Econometrics, v. 92, no. 2, p. 355–390. [Also available at https://doi.org/10.1016/S0304-4076(98)00095-5.]

Lee, C.J., Hirsch, R.M., Schwarz, G.E., Holtschlag, D.J., Preston, S.D., Crawford, C.G., and Vecchia, A.V., 2016, An evaluation of methods for estimating decadal stream loads: Journal of Hydrology, v. 542, p. 185–203, accessed March 13, 2018, at http://doi.org/10.1016/j.jhydrol.2016.08.059.

Press, W.H., Flannery, B.P., Teukolsky, S.A., and Vetterling, W.T., 1986, Numerical recipes: Cambridge, Cambridge University Press, 818 p.

Robert, C.P., and Casella, G., 1999, Monte Carlo statistical methods: New York, Springer, 507 p.

SAS Institute Inc., 2013, SAS/IML 12.3 user's guide: Cary, N.C., SAS Institute, Inc., 1,121 p.

Schnedler, W., 2005, Likelihood estimation for censored random vectors: Econometric Reviews, v. 24, no. 2, p. 195–217. [Also available at https://doi.org/10.1081/ETC-200067925.]

Schwarz, G.E., Hoos, A.B., Alexander, R.B., Smith, R.A., 2006, The SPARROW surface water-quality model—Theory, applications and user documentation: U.S. Geological Survey Techniques and Methods, book 6, chap. B3, 248 p. [Also available at https://doi.org/10.3133/tm6B3.]

Shenton, L.R., and Bowman, K.O., 1977, Maximum likelihood estimation in small samples: Oxford, Oxford University Press, 186 p.

Appendix 3. Derivation of Regularity Conditions Used To Evaluate the Covariance Matrix and Asymptotic Efficiency of the Estimates Produced by the Fluxmaster-K Algorithm

This appendix uses concepts developed in Schendler (2005) to establish two main results referenced in appendix 2: that the multivariate likelihood function of dependent censored variables retains the regularity conditions of a likelihood function for uncensored variables, and the maximum likelihood estimates are asymptotically efficient.

Regularity Conditions

The visibility state s defines a unique pattern of censored and uncensored outcomes among an $N \times 1$ vector $y = \{y_1, \dots, y_N\}'$ of the dependent variable of the log-likelihood function. W_s represents the range of y consistent with state s. Thus, if $N = 2$ there are four possible states: $s = 0$, in which both observations are censored and $W_0 = \{y_1 \le \bar{y}_1, y_2 \le \bar{y}_2\}$; $s = 1$, where y_1 is censored and y_2 is uncensored, making $W_1 = \{y_1 \le \bar{y}_1, y_2 > \bar{y}_2\}$; $s = 2$, where y_1 is uncensored and y_2 is censored, making $W_2 = \{y_1 > \bar{y}_1, y_2 \le \bar{y}_2\}$; and $s = 3$, where both values of y are uncensored and $W_3 = \{y_1 > \bar{y}_1, y_2 > \bar{y}_2\}$. With N observations this implies N^2 possible states. Note that the W_s sets are disjoint in the N-tuple y vector and that their joint union over all states corresponds to R^N, the entire space of y.

A visibility operator v^s is defined such that its application $v^s y$ to the dependent variable vector selects just the elements of y that are uncensored in state s. Accordingly, the complement operator $\bar{v}^s$ selects only elements of y that are censored in state s. Assume the probability density $p(y, \theta)$ for y to be a continuous function of the parameter vector $\theta \in \Theta$, where Θ is compact. The likelihood function for some state, s, is given by

$$\tilde{p}_s(v^s y|\theta) = \int_{W_s} p(y|\theta) d\mu(\bar{v}^s y)\,, \tag{3.1}$$

where $\ln \tilde{p}_s(v^s y|\theta)$ corresponds to the log-likelihood in equation 2.4, for some state s consistent with the visibility of the observed sample. Because the W_s ranges are disjointed across states, and their joint union comprises the entire support for y, the expected value of $\tilde{p}_s(v^s y|\theta)$ over all N^2 states is

$$\sum_{s=0}^{N^2-1} \int_{W_s} \tilde{p}_s(v^s y|\theta) d\mu(v^s y) = \sum_{s=0}^{N^2-1} \int_{W_s} \int_{W_s} p(y|\theta) d\mu(\bar{v}^s y)\, d\mu(v^s y)$$

$$= \int p(y|\theta) d\mu(y) = 1. \tag{3.2}$$

The probability of state s, is given by

$$\Psi_s = \int_{W_s} p(y|\theta) d\mu(y) . \tag{3.3}$$

Define $p_s(v^s y|\theta)$ to be the probability density of the uncensored observations in state s, given that state s has occurred. This probability can be expressed as

$$p_s(v^s y|\theta) = \int_{W_s} p(y|\theta) d\mu(\bar{v}^s y) / \Psi_s . \tag{3.4}$$

Under the assumption that $p(y|\theta)$ adheres to regularity conditions, then the following relations apply:

$$\frac{\partial}{\partial \theta} \int p(y|\theta) d\mu(y) = \int p_\theta(y|\theta) d\mu(y) , \tag{3.5}$$

$$\frac{\partial^2}{\partial \theta \partial \theta'} \int p(y|\theta) d\mu(y) = \int p_{\theta\theta'}(y|\theta) d\mu(y) , \tag{3.6}$$

and

$$\frac{\partial^3}{\partial \theta_a \partial \theta_b \partial \theta_c} \int p(y|\theta) d\mu(y) = \int p_{\theta_a \theta_b \theta_c}(y|\theta) d\mu(y) . \tag{3.7}$$

The expectation of the partial derivative L_θ, evaluated at the true coefficient values θ_0, is given by

$$\begin{aligned} E[L_\theta]|_{\theta=\theta_0} &= \sum_{s=0}^{N^2-1} \Psi_s \int_{W_s} \frac{\tilde{p}_{s\theta}(v^s y|\theta_0)}{\tilde{p}_s(v^s y|\theta_0)} p_s(v^s y|\theta_0) d\mu(v^s y) \\ &= \sum_{s=0}^{N^2-1} \Psi_s \int_{W_s} \frac{\tilde{p}_{s\theta}(v^s y|\theta_0)}{\tilde{p}_s(v^s y|\theta_0)} \frac{\tilde{p}_s(v^s y|\theta_0)}{\Psi_s} d\mu(v^s y) \\ &= \sum_{s=0}^{N^2-1} \int_{W_s} \tilde{p}_{s\theta}(v^s y, \theta_0) d\mu(v^s y) \\ &= \frac{\partial}{\partial \theta} \underbrace{\int p(y|\theta_0) d\mu(y)}_{=1} = 0, \end{aligned} \tag{3.8}$$

where the first equality evaluates the expectation based on $p_s(v^s y|\theta)$, the probability of the uncensored values of y given state s has occurred, times Ψ_s, the probability of state s; the second equality follows because $\tilde{p}_s(v^s y|\theta) = \Psi_s p_s(v^s y|\theta)$ (substitute eq. 3.1 into eq. 3.4); the third equality cancels common terms; and the fourth equality follows because the W_s ranges are disjoint, cover the entire range of y, and are not a function of θ, and because of the regularity condition for $p(y|\theta)$ (see eq. 3.5).

Similarly, for the expectation of the second-order partial derivative of the log-likelihood, has the equivalent relation

$$E[L_{\theta\theta'}]|_{\theta=\theta_0} =$$

$$\sum_{s=0}^{N^2-1} \Psi_s \int_{W_s} \left(\frac{\tilde{p}_{s\theta\theta'}(v^s y|\theta_0)}{\tilde{p}_s(v^s y|\theta_0)} - \frac{\tilde{p}_{s\theta}(v^s y|\theta_0)\tilde{p}_{s\theta'}(v^s y|\theta_0)}{\tilde{p}_s(v^s y|\theta_0)^2} \right) p_s(v^s y|\theta) d\mu(v^s y)$$

$$= \sum_{s=0}^{N^2-1} \int_{W_s} \tilde{p}_{s\theta\theta'}(v^s y, \theta_0) d\mu(v^s y) - E[L_\theta L_{\theta'}]|_{\theta=\theta_0}$$

$$= \frac{\partial^2}{\partial\theta\partial\theta'} \underbrace{\int p(y|\theta) d\mu(y)}_{=1} - E[L_\theta L_{\theta'}]|_{\theta=\theta_0} = -E[L_\theta L_{\theta'}]|_{\theta=\theta_0}. \qquad (3.9)$$

The third-order partial derivative of the log-likelihood function, expressed as $\partial^3 L/\partial\theta_a\partial\theta_b\partial\theta_c \equiv L_{abc}$ and evaluated at the true values of the parameters, is given by

$$E[L_{abc}]|_{\theta=\theta_0} = \sum_{s=0}^{N^2-1} \Psi_s \int_{W_s} \left\{ \frac{\tilde{p}_{sabc}(v^s y|\theta_0)}{\tilde{p}_s(v^s y|\theta_0)} - \frac{\tilde{p}_{sab}(v^s y|\theta_0)\tilde{p}_{sc}(v^s y|\theta_0)}{\tilde{p}_s(v^s y|\theta_0)^2} \right.$$

$$- \frac{\tilde{p}_{sac}(v^s y|\theta_0)\tilde{p}_{sb}(v^s y|\theta_0)}{\tilde{p}_s(v^s y|\theta_0)^2} - \frac{\tilde{p}_{sbc}(v^s y|\theta_0)\tilde{p}_{sa}(v^s y|\theta_0)}{\tilde{p}_s(v^s y|\theta_0)^2}$$

$$\left. + 2\frac{\tilde{p}_{sa}(v^s y|\theta_0)\tilde{p}_{sb}(v^s y|\theta_0)\tilde{p}_{sc}(v^s y|\theta_0)}{\tilde{p}_s(v^s y|\theta_0)^3} \right\} p_s(v^s y|\theta_0) d\mu(v^s y)$$

$$= \underbrace{\sum_{s=0}^{N^2-1} \int_{W_s} \tilde{p}_{sabc}(v^s y|\theta_0) d\mu(v^s y)}_{=0} - \{E[(L_{ab} + L_a L_b)L_c]$$

$$-E[(L_{ac} + L_a L_c)L_b] - E[(L_{bc} + L_b L_c)L_a] + 2E[L_a L_b L_c]\}|_{\theta=\theta_0}$$

$$= -\{E[L_{ab}L_c] + E[L_{ac}L_b] + E[L_{bc}L_a] + E[L_a L_b L_c]\}|_{\theta=\theta_0}. \qquad (3.10)$$

Asymptotic Efficiency

The property of asymptotic efficiency for the maximum likelihood coefficient estimates is established following the approach of Schendler (2005). The first step is to derive the Cramer-Rao lower bound for some unbiased estimator $T(v^s y)$ that is based only on uncensored data. Because the estimator is unbiased, a valid expression for the true values of the coefficients is

$$\theta_0 = E[T(v^s y)] = \sum_{s=0}^{N^2-1} \Psi_s \int_{W_s} T(v^s y) p_s(v^s y|\theta) d\mu(v^s y)$$

$$= \sum_{s=0}^{N^2-1} \int_{W_s} T(v^s y) \tilde{p}_s(v^s y|\theta_0) d\mu(v^s y)\,. \qquad (3.11)$$

Taking the first-order partial derivative, we have

$$\begin{aligned}\frac{\partial \theta_0}{\partial \theta'} = I &= \sum_{s=0}^{N^2-1} \int_{W_s} T(v^s y)\tilde{p}_{s\theta'}(v^s y|\theta_0)d\mu(v^s y) \\ &= \sum_{s=0}^{N^2-1} \int_{W_s} T(v^s y)\frac{\tilde{p}_{s\theta'}(v^s y|\theta_0)}{\tilde{p}_s(v^s y|\theta_0)}\tilde{p}_s(v^s y|\theta_0)d\mu(v^s y) \\ &= \sum_{s=0}^{N^2-1} \Psi_s \int_{W_s} T(v^s y)\frac{\tilde{p}_{s\theta'}(v^s y|\theta_0)}{\tilde{p}_s(v^s y|\theta_0)}p_s(v^s y|\theta_0)d\mu(v^s y) \\ &= E\left[T(v^s y)\frac{\tilde{p}_{s\theta'}(v^s y|\theta_0)}{\tilde{p}_s(v^s y|\theta_0)}\right]. \end{aligned} \quad (3.12)$$

From equation 3.8, we have $E[\tilde{p}_{s\theta'}(v^s y|\theta_0)/\tilde{p}_s(v^s y|\theta_0)] = 0$, implying we can subtract the term $\theta_0 E[\tilde{p}_{s\theta'}(v^s y|\theta_0)/\tilde{p}_s(v^s y|\theta_0)]$ from equation 3.12 without changing the result. Therefore, the covariance between $T(v^s y)$ and $L_{\theta'}$ is given by

$$I = E\left[(T(v^s y) - \theta_0)\frac{\tilde{p}_{s\theta'}(v^s y|\theta_0)}{\tilde{p}_s(v^s y|\theta_0)}\right] = E[(T(v^s y) - \theta_0)L_{\theta'}]. \quad (3.13)$$

Define $\mathcal{T} = T(v^s y) - \theta_0$ to be the error in the unbiased estimate. As shown in equation 3.8, $E[L_{\theta'}]|_{\theta=\theta_0} = 0$, and because $\mathcal{T}$ is unbiased, implying $E[\mathcal{T}] = 0$, then both the $\mathcal{T}$ and L_θ statistics have a mean of 0. Therefore, the covariance matrix between $\mathcal{T}$ and L_θ takes the form

$$COV(\mathcal{T}, L_\theta) = \begin{bmatrix} E[\mathcal{T}\mathcal{T}'] & E[\mathcal{T}L_{\theta'}]|_{\theta=\theta_0} \\ E[\mathcal{T}L_{\theta'}]|'_{\theta=\theta_0} & E[L_\theta L_{\theta'}]|_{\theta=\theta_0} \end{bmatrix} = \begin{bmatrix} E[\mathcal{T}\mathcal{T}'] & I \\ I & E[L_\theta L_{\theta'}]|_{\theta=\theta_0} \end{bmatrix}, \quad (3.14)$$

where $E[\mathcal{T}\mathcal{T}']$ is the covariance matrix for $T(v^s y)$, and the second equality is due to equation 3.13. Equation 3.14 is a valid covariance matrix, implying that $COV(\mathcal{T}, L_\theta)$ is positive semidefinite. Therefore, a quadratic form created from the row vector $[x' \quad -x'(E[L_\theta L_{\theta'}])^{-1}]$, for any vector x, is given by

$$\begin{aligned}&\left[x' \quad -x'\left(E[L_\theta L_{\theta'}]|_{\theta=\theta_0}\right)^{-1}\right]\begin{bmatrix} E[\mathcal{T}\mathcal{T}'] & I \\ I & E[L_\theta L_{\theta'}]|_{\theta=\theta_0} \end{bmatrix}\begin{bmatrix} x \\ -\left(E[L_\theta L_{\theta'}]|_{\theta=\theta_0}\right)^{-1}x \end{bmatrix} \\ &= x'\left(E[\mathcal{T}\mathcal{T}'] - \left(E[L_\theta L_{\theta'}]|_{\theta=\theta_0}\right)^{-1}\right)x \geq 0, \end{aligned} \quad (3.15)$$

where the inequality is a consequence of the positive semidefiniteness of $COV(\mathcal{T}, L_\theta)$.

Thus, $\left(E[L_\theta L_{\theta'}]|_{\theta=\theta_0}\right)^{-1}$ represents the Cramer-Rao lower bound on the covariance matrix of any unbiased estimator of θ_0. If the unbiased estimator $T(v^s y)$ is consistent, then its covariance matrix degenerates to the zero matrix as sample size approaches infinity, as does the covariance matrix $(E[L_\theta L_{\theta'}])^{-1}$. These results can be trivially generalized to the case where the covariance between $\sqrt{N}\mathcal{T}$ and $L_{\theta'}/\sqrt{N}$ is formed, giving the Cramer-Rao lower bound for the asymptotic covariance matrix

$$x'\left(\lim_{N\to\infty} COV\left(\sqrt{N}\mathcal{T}\right) - \left(\lim_{N\to\infty} N^{-1}E[L_\theta L_{\theta'}]|_{\theta=\theta_0}\right)^{-1}\right)x \geq 0. \quad (3.16)$$

Aside from being the asymptotic Cramer-Rao lower bound, as shown in equation 3.16, $\left(\lim_{N\to\infty} N^{-1}E[L_\theta L_{\theta'}]|_{\theta=\theta_0}\right)^{-1}$ is also the asymptotic covariance matrix for the maximum likelihood estimator. Therefore, the maximum likelihood estimator is asymptotically efficient.

Reference Cited

Schendler, W., 2005, Likelihood estimation for censored random vectors: Econometric Reviews, v. 24, no. 2, p. 195-217. [Also available at https://doi.org/10.1081/ETC-200067925.]

For more information about this report, contact:
Director, Upper Midwest Water Science Center
U.S. Geological Survey
8505 Research Way
Middleton, WI 53562
dc_wi@usgs.gov
or visit our website at
https://www.usgs.gov/centers/umid-water

Publishing support provided by the Pembroke and Madison Publishing Service Centers.

ISSN 2328-0328 (online)
https://doi.org/10.3133/sir20195069