National Water-Quality Assessment Program

Development and Application of Regression Models for Estimating Nutrient Concentrations in Streams of the Conterminous United States, 1992–2001

Scientific Investigations Report 2009–5199

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

Development and Application of Regression Models for Estimating Nutrient Concentrations in Streams of the Conterminous United States, 1992–2001

By Norman E. Spahr, David K. Mueller, David M. Wolock, Kerie J. Hitt, and JoAnn M. Gronberg

National Water-Quality Assessment Program

Scientific Investigations Report 2009–5199

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

U.S. Department of the Interior
KEN SALAZAR, Secretary

U.S. Geological Survey
Marcia K. McNutt , Director

U.S. Geological Survey, Reston, Virginia: 2010

For product and ordering information:
World Wide Web: http://www.usgs.gov/pubprod
Telephone: 1-888-ASK-USGS

For more information on the USGS--the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment:
World Wide Web: http://www.usgs.gov
Telephone: 1-888-ASK-USGS

Suggested citation:
Spahr, N.E., Mueller, D.K., and Wolock, D.W., 2010, Development and application of regression models for estimating nutrient concentrations in streams of the conterminous United States, 1992–2001: U.S. Geological Survey Scientific Investigations Report 2009-5199, 22 p.

Foreword

The U.S. Geological Survey (USGS) is committed to providing the Nation with reliable scientific information that helps to enhance and protect the overall quality of life and that facilitates effective management of water, biological, energy, and mineral resources (http://www.usgs.gov/). Information on the Nation's water resources is critical to ensuring long-term availability of water that is safe for drinking and recreation and is suitable for industry, irrigation, and fish and wildlife. Population growth and increasing demands for water make the availability of that water, measured in terms of quantity and quality, even more essential to the long-term sustainability of our communities and ecosystems.

The USGS implemented the National Water-Quality Assessment (NAWQA) Program in 1991 to support national, regional, State, and local information needs and decisions related to water-quality management and policy (http://water.usgs.gov/nawqa). The NAWQA Program is designed to answer: What is the quality of our Nation's streams and ground water? How are conditions changing over time? How do natural features and human activities affect the quality of streams and ground water, and where are those effects most pronounced? By combining information on water chemistry, physical characteristics, stream habitat, and aquatic life, the NAWQA Program aims to provide science-based insights for current and emerging water issues and priorities. From 1991 to 2001, the NAWQA Program completed interdisciplinary assessments and established a baseline understanding of water-quality conditions in 51 of the Nation's river basins and aquifers, referred to as Study Units (http://water.usgs.gov/nawqa/studyu.html).

National and regional assessments are ongoing in the second decade (2001–2012) of the NAWQA Program as 42 of the 51 Study Units are selectively reassessed. These assessments extend the findings in the Study Units by determining water-quality status and trends at sites that have been consistently monitored for more than a decade, and filling critical gaps in characterizing the quality of surface water and ground water. For example, increased emphasis has been placed on assessing the quality of source water and finished water associated with many of the Nation's largest community water systems. During the second decade, NAWQA is addressing five national priority topics that build an understanding of how natural features and human activities affect water quality, and establish links between sources of contaminants, the transport of those contaminants through the hydrologic system, and the potential effects of contaminants on humans and aquatic ecosystems. Included are studies on the fate of agricultural chemicals, effects of urbanization on stream ecosystems, bioaccumulation of mercury in stream ecosystems, effects of nutrient enrichment on aquatic ecosystems, and transport of contaminants to public-supply wells. In addition, national syntheses of information on pesticides, volatile organic compounds (VOCs), nutrients, trace elements, and aquatic ecology are continuing.

The USGS aims to disseminate credible, timely, and relevant science information to address practical and effective water-resource management and strategies that protect and restore water quality. We hope this NAWQA publication will provide you with insights and information to meet your needs, and will foster increased citizen awareness and involvement in the protection and restoration of our Nation's waters.

The USGS recognizes that a national assessment by a single program cannot address all water-resource issues of interest. External coordination at all levels is critical for cost-effective management, regulation, and conservation of our Nation's water resources. The NAWQA Program, therefore, depends on advice and information from other agencies—Federal, State, regional, interstate, Tribal, and local—as well as nongovernmental organizations, industry, academia, and other stakeholder groups. Your assistance and suggestions are greatly appreciated.

Matthew C. Larsen
Associate Director for Water

Contents

Figures

Tables

Conversion Factors

[Inch/Pound to SI]

Multiply	By	To obtain
	Length	
inch (in.)	2.54	centimeter (cm)
inch (in.)	25.4	millimeter (mm)
foot (ft)	0.3048	meter (m)
mile (mi)	1.609	kilometer (km)
	Area	
acre	4,047	square meter (m^2)
acre	0.4047	hectare (ha)
square mile (mi^2)	259.0	hectare (ha)
square mile (mi^2)	2.590	square kilometer (km^2)
	Volume	
ounce, fluid (fl. oz)	0.02957	liter (L)
pint (pt)	0.4732	iter (L)
quart (qt)	0.9464	liter (L)
gallon (gal)	3.785	liter (L)
gallon (gal)	0.003785	cubic meter (m^3)
cubic inch (in^3)	0.01639	liter (L)
cubic foot (ft^3)	0.02832	cubic meter (m^3)
acre-foot (acre-ft)	1,233	cubic meter (m^3)
	Flow rate	
acre-foot per day (acre-ft/d)	0.01427	cubic meter per second (m^3/s)
cubic foot per second (ft^3/s)	0.02832	cubic meter per second (m^3/s)
	Mass	
ounce, avoirdupois (oz)	28.35	gram (g)
pound, avoirdupois (lb)	0.4536	kilogram (kg)
	Application rate	
pounds per acre per year [(lb/acre)/yr]	1.121	kilograms per hectare per year [(kg/ha)/yr]

Temperature in degrees Celsius (°C) may be converted to degrees Fahrenheit (°F) as follows:

$$°F=(1.8\times°C)+32$$

Temperature in degrees Fahrenheit (°F) may be converted to degrees Celsius (°C) as follows:

$$°C=(°F-32)/1.8$$

Vertical coordinate information is referenced to the "North American Vertical Datum of 1988 (NAVD 88)."

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

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

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

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

Development and Application of Regression Models for Estimating Nutrient Concentrations in Streams of the Conterminous United States, 1992–2001

By Norman E. Spahr, David K. Mueller, David M. Wolock, Kerie J. Hitt, and JoAnn M. Gronberg

Abstract

Data collected for the U.S. Geological Survey National Water-Quality Assessment program from 1992–2001 were used to investigate the relations between nutrient concentrations and nutrient sources, hydrology, and basin characteristics. Regression models were developed to estimate annual flow-weighted concentrations of total nitrogen and total phosphorus using explanatory variables derived from currently available national ancillary data. Different total-nitrogen regression models were used for agricultural (25 percent or more of basin area classified as agricultural land use) and nonagricultural basins. Atmospheric, fertilizer, and manure inputs of nitrogen, percent sand in soil, subsurface drainage, overland flow, mean annual precipitation, and percent undeveloped area were significant variables in the agricultural basin total nitrogen model. Significant explanatory variables in the nonagricultural total nitrogen model were total nonpoint-source nitrogen input (sum of nitrogen from manure, fertilizer, and atmospheric deposition), population density, mean annual runoff, and percent base flow.

The concentrations of nutrients derived from regression (CONDOR) models were applied to drainage basins associated with the U.S. Environmental Protection Agency (USEPA) River Reach File (RF1) to predict flow-weighted mean annual total nitrogen concentrations for the conterminous United States. The majority of stream miles in the Nation have predicted concentrations less than 5 milligrams per liter. Concentrations greater than 5 milligrams per liter were predicted for a broad area extending from Ohio to eastern Nebraska, areas spatially associated with greater application of fertilizer and manure. Probabilities that mean annual total-nitrogen concentrations exceed the USEPA regional nutrient criteria were determined by incorporating model prediction uncertainty. In all nutrient regions where criteria have been established, there is at least a 50 percent probability of exceeding the criteria in more than half of the stream miles.

Dividing calibration sites into agricultural and nonagricultural groups did not improve the explanatory capability for total phosphorus models. The group of explanatory variables that yielded the lowest model error for mean annual total phosphorus concentrations includes phosphorus input from manure, population density, amounts of range land and forest land, percent sand in soil, and percent base flow. However, the large unexplained variability and associated model error precluded the use of the total phosphorus model for nationwide extrapolations.

Introduction

During the past decade, nutrients have consistently been cited as one of the leading pollutants in rivers and streams (see, for example, U.S. Environmental Protection Agency, 2002b and 2007). Nutrients are chemical compounds that contain nitrogen (N) or phosphorus (P). These compounds are essential to plant and animal nutrition, but in high concentrations they can be contaminants in water. Nutrient compounds are affected by chemical, biological, and hydrologic processes that can change their form and can transfer them to or from water, soil, biological organisms, and the atmosphere.

In 1991 the U.S. Geological Survey (USGS) began full implementation of the National Water-Quality Assessment (NAWQA) program to assess the status and trends of the Nation's surface and ground waters. The design of the NAWQA program is presented in Gilliom and others (1995). Nutrients are one of the primary synthesis topics of the NAWQA program.

Nutrients were investigated in streams and rivers sampled as part of the U.S. Geological Survey National Water-Quality Assessment (NAWQA) Program. Nutrient data were collected in 20 NAWQA study units during 1992–95, 16 study units during 1996-98, and 15 study units during 1999–2001. Mueller and Spahr (2005) compiled nutrient data collected during Cycle I (1992–2001) of the NAWQA program. To facilitate comparisons among sampling sites with variable sampling frequency, Mueller and Spahr (2006) computed daily loads by using regression models that relate constituent transport to streamflow and time. The specific method used for model calibration and load estimation is described by Runkel and others (2004). Mueller and Spahr (2006, page 6) used the model results to compute mean annual loads, yields, and concentrations of total nitrogen and total phosphorus for NAWQA sampling sites. Sufficient data were available to calculate annual concentrations at all 481 sites for total nitrogen and at 463 sites for total phosphorus.

Purpose and Scope

A national comparison of nutrient conditions among stream sites was facilitated by modeling annual loads and concentrations (Mueller and Spahr, 2006). Nutrient concentrations, yields, and loads were compared to basin land-use characteristics and major nonpoint sources of nitrogen and phosphorus. This report extends the previous analysis and extrapolates results to basins where water-quality samples were not collected during Cycle I of the NAWQA program by development and application of regression models for estimating nutrient concentration in streams.

Methods

The flow-weighted mean concentrations of total nitrogen and total phosphorus as reported by Mueller and Spahr (2005 and 2006) are considered to be the "observed" concentrations in this report. These concentrations are used in correlation and regression analyses, along with various basin characteristics, in order to estimate mean annual total nitrogen and phosphorus concentrations in streams throughout the conterminous United States.

Basin Characteristics

Characteristics of the drainage basins upstream from NAWQA sites were derived from data sets available for the entire conterminous United States. Atmospheric and manure nonpoint-source nutrient inputs were based on county level estimates as described by Ruddy and others (2006). Fertilizer nonpoint-source inputs were based on methodology described by Ruddy and others (2006) using updated data (J.M. Gronberg, U.S. Geological Survey, written commun., 2008). Population density was derived from the 1990 census and block group data (U.S. Bureau of the Census, 1991a and 1991b).

Percent land use cover was determined from the 1992 National Land Cover Data (NLCD) digital maps as revised by Nakagaki and Wolock (2005). Agriculture land use is the combined percentages of NLCD classes' row crop, small grains, fallow, orchards and vineyard, and pasture and hay (NLCD classes 61, 62, 81, 82, 83, and 84). Undeveloped land use includes percentages of grasslands, shrub lands, all forested lands, as well as, open water, wetlands, tundra, and bare rock (NLCD classes 11, 12, 31, 32, 33, 41, 42, 43, 51, 71, 72, 91, and 92).

Percent sand in soil was obtained from the U.S. Department of Agriculture's State Soil Geographic database (Natural Resources Conservation Service, 1994). Mean annual runoff (1950–1981) was determined from a map by Gebert and others (1987). Base flow index and Horton overland flow were obtained from online data sets (Wolock, 2003a and 2003b). Information on tile drains is from the National Resources Conservation Service (1995). Precipitation and evaporation (1961–90) are derived from a model developed by Daly and others (2001)

Correlation and Regression Analysis

Nonparametric Spearman coefficients (Helsel and Hirsch, 1993) were used to examine the various correlations between flow-weighted concentrations and basin characteristics. Results of the correlation analysis were used to select potential explanatory variables for regression models. During the correlation and regression calibration phase, nonpoint nutrient input terms (fertilizer, manure, and atmospheric deposition) represent the NAWQA high-intensity period for each sampling site as described by Mueller and Spahr (2005, ancillary data section). Other terms represent longer time periods as described in the section titled "Basin Characteristics."

The multiple linear regression models were built in a stepwise process. The dependent variable in each model was the base-10 logarithm ($\log_{10}$) of the flow-weighted concentration. Initial explanatory variables included the various nonpoint-source inputs. Correlations between the residuals from these initial models and additional basin characteristics led to the selection of subsequent explanatory variables. In some cases, transformations (logarithmic, square root, and fourth root) of the explanatory variables were employed to achieve linearity, normality of residuals, and homogeneity of residual variance (Helsel and Hirsch, 1993). Partial residual plots were used to determine if selected transformations of explanatory variables were appropriate. With the addition of each new variable the adjusted R squared and the root mean square error were examined to evaluate the potential improvement gained by the addition of another variable. Explanatory variables were added to the model until no tangible improvements were seen or until subsequent variables were not significant at an alpha level of 0.05. Although somewhat subjective, this method balances model simplicity with the ability of the model to explain the variability of nutrient concentrations. All statistics, correlations, and regressions were analyzed using SAS statistical software (SAS Institute Inc., 1999).

To determine if the significance and form of a set of explanatory variables was dependent on a specific set of stations, one-third of the stations were withheld from model calibration and the regression coefficients were determined using the remaining two-thirds of the stations. This exercise was replicated 200 times, and the significance and form of each coefficient were compared among the calibrated models. If an explanatory variable was not significant in a large percentage of these models or its coefficient was positive in some models and negative in others, then that variable was removed from consideration.

Nationwide Prediction

The U.S. Environmental Protection Agency (USEPA) River Reach File version 1.0 (RF1), described by Nolan and others (2003) was used as the basis for predicting nutrient concentrations throughout the conterminous United States. RF1 is a georeferenced database of stream reaches that comprise the surface water drainage of the United States. Basin characteristics and nonpoint-source nutrient inputs were determined for basins upstream from each RF1 node (the point representing the downstream end of each reach). The previously calibrated regression equations then were used to predict concentrations at each node point. Retransformation bias was corrected using the smearing estimate of Duan (1983) as outlined by Helsel and Hirsch (1993).

Predicted concentrations were assumed to apply to the entire RF1 reach upstream from the node. Similar methodology was used in the NAWQA program to predict atrazine concentrations in stream water, as documented by Gilliom and others (2006), and dieldrin concentrations in fish tissue, as documented by Nowell and others (2006). In addition to predicting nutrient concentrations for each RF1 reach, the probability of a concentration being greater than some reference value can be determined. Because residuals from the regression equations are normally distributed, the probability of an estimated concentration being greater than some value, *X*, can be represented by the area under a normal curve above the value *Z*:

$$Z = \frac{C - X}{S_c} \qquad (1)$$

where

- Z is a standard normal deviate,
- C is the concentration estimated from the regression model,
- X is the reference concentration, and
- S_C is the standard error of C, equal to the prediction error of the regression model.

The primary nonpoint source inputs (fertilizer, manure, and atmospheric deposition) represent the entire NAWQA Cycle I period (1992–2001) for nationwide simulation. Therefore, the simulated concentrations represent the cycle one period of the NAWQA program.

Correlation between Nutrient Concentrations and Basin Characteristics

Nutrient concentrations in stream water vary with upstream land-use practices that affect nutrient sources and transport of nutrients to stream channels. On a national scale there is great variability associated with stream nutrient concentrations and related basin characteristics. Spearman correlation coefficients between flow-weighted concentrations and various explanatory variables are listed in table 1. In addition to analyzing all sites together, sites were classified as agricultural (agricultural land use greater than or equal to 25 percent of the basin area) or nonagricultural (fig. 1).

For some explanatory variables, such as percent sand in soil and percent area with subsurface drains, correlation with nutrient concentrations is greater in agricultural basins than in nonagricultural basins, indicating nutrient concentrations in streams might be responding to an interaction between land use and other basin characteristics. Nonpoint-source inputs of nitrogen and phosphorus from fertilizer are more highly correlated with stream nutrient concentrations in agricultural areas than nonagricultural areas, reflecting the greater application associated with agricultural land use. Population density is a surrogate for point sources of nutrients. Correlations between population density and stream nutrient concentrations are greater in nonagricultural areas which include many urban watersheds.

Mueller and Spahr (2006) have shown that concentrations of all nutrients in streams increase as development in the basin increases. Percent of undeveloped land use has relatively high negative correlations with nutrient concentrations (table 1). Less development and lower percentage of agricultural land use corresponds with lower concentrations, and the magnitude of the correlation is greater for the agricultural than nonagricultural sites. Peterson and others (2004) describe higher phosphorus concentrations in basins with higher percentage of rangeland for NAWQA sites in Wyoming and this also appears to be the case for the national set of nonagricultural basins in this study.

Development of Models for Total Nitrogen

An initial regression model relating nitrogen concentrations to non-point sources (atmospheric, fertilizer, and manure nitrogen) for all sites explained only about 53 percent of the variability. Various basin characteristics were added to the all sites model based upon the correlation of residuals with additional ancillary data. Results from these regression models could not adequately explain the variability associated with different land use types and these models consistently underestimated concentrations for many agricultural basins. Finally,

Table 1. Correlation between flow-weighted annual mean total-nitrogen or total-phosphorus concentrations at National Water-Quality Assessment sampling sites and selected characteristics of the upstream basin.

[n, number of samples; kg/km^2, kilograms per square kilometer; km^2, square kilometer; bold type indicates absolute value of correlation is greater than 0.5; --, not applicable]

Explanatory variable	Spearman correlation coefficient					
	All Sites		Agricultural Sites		Nonagricultural Sites	
	Total nitrogen n=462	Total phosphorus n=440	Total nitrogen n=220	Total phosphorus n=214	Total nitrogen n=242	Total phosphorus n=226
Atmospheric nitrogen input, kg/km^2	0.35	--	0.30	--	0.20	--
Fertilizer nitrogen input, kg/km^2	**0.71**	--	**0.62**	--	**0.54**	--
Manure nitrogen input, kg/km^2	0.49	--	0.28	--	0.19	--
Fertilizer phosphorus input, kg/km^2	--	0.44	--	0.35	--	0.26
Manure phosphorus input, kg/km^2	--	0.34	--	0.22	--	0.15
Population density, per km^2	0.30	0.20	0.04	–0.05	0.50	0.29
Undeveloped land use, percent of basin area	**–0.74**	**–0.50**	**–0.73**	–0.50	**–0.57**	–0.30
Forest land use, percent of basin area	**–0.69**	**–0.63**	**–0.64**	–0.47	**–0.50**	**–0.62**
Range land use, percent of basin area	–0.09	–0.13	0.13	0.17	0.05	0.35
Sand in soil, percent	–0.38	–0.31	–0.32	–0.38	–0.11	–0.06
Subsurface drainage, percent of basin area	0.39	0.22	0.48	0.23	0.12	–0.01
Horton overland flow, percent of long-term mean streamflow	0.25	0.34	0.00	0.35	0.16	0.17
Base flow index, percent of long-term mean streamflow	–0.38	–0.36	0.00	–0.25	–0.35	–0.20
Mean annual runoff 1951–90, inches	–0.36	–0.39	–0.24	–0.15	–0.35	–0.47
Mean annual precipitation, inches	–0.11	–0.14	–0.34	–0.10	–0.08	–0.26
Mean annual evaporation, inches	0.16	0.36	–0.09	0.18	0.26	0.46

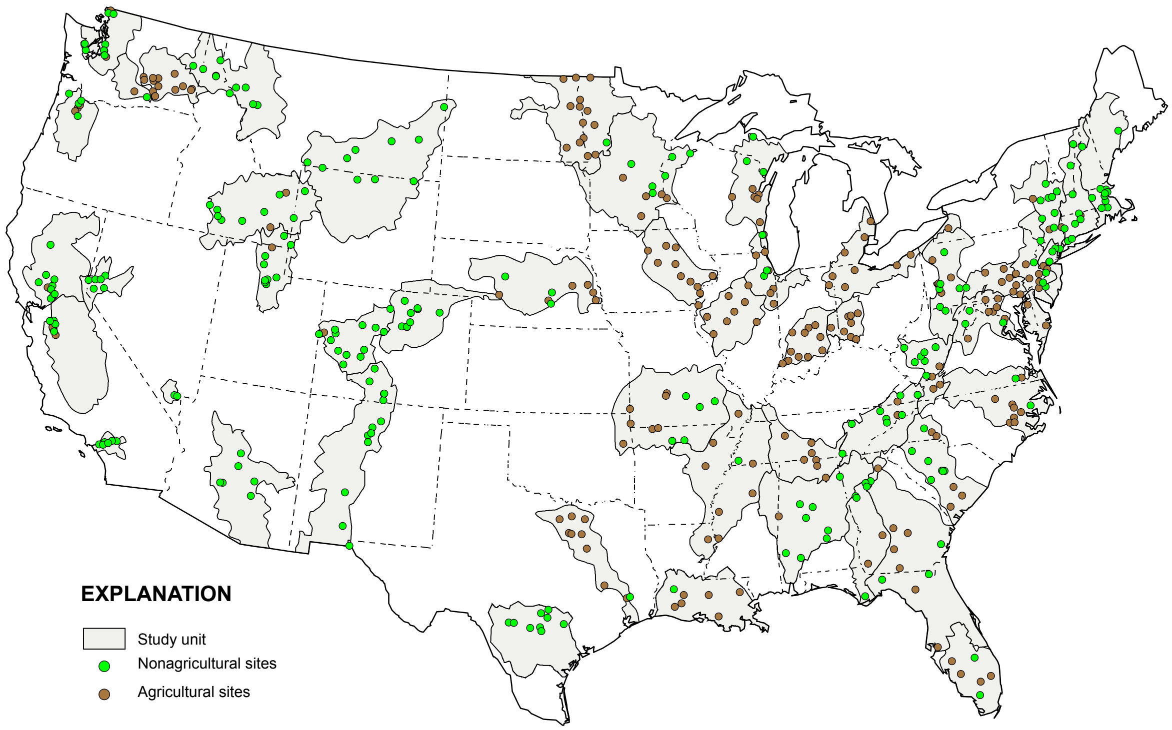


Figure 1. Location of National Water-Quality Assessment sampling sites downstream from agricultural and nonagricultural areas, as classified for use in the development of total-nitrogen regression models.

sub-setting sites into agricultural and nonagricultural groups led to regression models with acceptable explanatory capabilities. Sites where agricultural land use was at least 25 percent of the upstream basin area were considered agricultural. The geographic distribution of 220 agricultural sites and 242 nonagricultural sites, based on this classification, is shown in figure 1. The variation in spatial density is a result of regional differences in arable land and also is influenced by the design of the NAWQA sampling network (Gilliom and others, 1995).

Agricultural Sites

Explanatory variables used in the concentrations of nutrients derived from regression (CONDOR) model for total nitrogen downstream from agricultural areas are listed in table 2. Regression coefficients for these variables were significant (p-value < 0.01) in more than 92 percent of the 200 simulation tests where random selections of two-thirds of the sites were used for calibration of the regression model. Atmospheric nitrogen input and subsurface drainage were the variables that most frequently became nonsignificant in the tests. For the model calibrated using data from all agricultural sites, the root-mean-square error is 0.19, the adjusted R-squared value is 0.71, and the bias-correction factor used for de-transforming estimated total-nitrogen concentrations is 1.10.

As expected, all sources of nitrogen (atmospheric, fertilizer, and manure) in the model are positively correlated with total-nitrogen concentrations in streams. In general, total-nitrogen concentrations are higher downstream from areas with higher rates of atmospheric deposition, fertilizer application, or manure accumulation. The presence of subsurface drainage also is positively correlated with total-nitrogen concentrations, probably because installation of drainage tile short-circuits the normal flow path and results in rapid movement of nitrogen to streams. The remaining explanatory variables in the model have negative coefficients; higher values of these basin characteristics generally are associated with lower total-nitrogen concentrations. Percent sand in soils is an indication of potential infiltration. Where percent sand is high, infiltration also is likely to be high, and there will be less direct runoff, and therefore less nitrogen transport, to streams. The negative coefficient on Horton overland flow seems counter-intuitive, because greater overland flow could result in greater nitrogen transport. Overland flow occurs when precipitation exceeds infiltration capacity, and this is more likely on land that is less suitable for intensive agriculture, either because of soil properties or slope. Thus, the overland-flow variable in this model could be associated with basins where sources of nitrogen are limited, particularly in proximity to stream channels. Precipitation creating overland flow from areas with limited sources of nitrogen may have a dilution effect. The negative coefficient on precipitation indicates that areas with higher annual precipitation tend to have somewhat lower total nitrogen concentrations. This is hypothesized as a dilution effect. Runoff and streamflow were also tested as explanatory

Table 2. Explanatory variables in the model of $\log_{10}$ flow-weighted mean annual total-nitrogen concentrations downstream from agricultural areas.

[kg/km^2, kilograms per square kilometer]

Explanatory variable (transformation)	Regression coefficient	Standard error	p-value
Intercept	0.53537	0.16889	0.0021
Atmospheric nitrogen input, kg/km^2 (square root)	0.01437	0.00402	0.0004
Fertilizer nitrogen input, kg/km^2 (square root)	0.00429	0.00114	0.0002
Manure nitrogen input, kg/km^2 (fourth root)	0.06193	0.00876	<0.0001
Sand in soil, percent (fourth root)	–0.26667	0.04702	<0.0001
Subsurface drainage (1 if more than 5 percent of basin area, 0 otherwise)	0.14511	0.03897	0.0003
Horton overland flow, percent of long-term mean streamflow	–0.01408	0.00310	<0.0001
Undeveloped land use, percent of basin area	–0.00518	0.00112	<0.0001
Mean annual precipitation, inches	–0.00536	0.00120	<0.0001

variables and typically had negative coefficients, but precipitation was a more consistent predictor variable. Finally, the amount of undeveloped land use is negatively correlated with total-nitrogen concentrations. In basins with greater development (agricultural and urban areas), sources of nitrogen are more common and the probability of direct transport to stream channels is greater.

A comparison of the observed and estimated total nitrogen at agricultural sites is shown in figure 2. The regression model generally performs well throughout the range of total nitrogen concentrations. Three sites (the Palouse River in Washington, Sacramento Slough in California, and the Niangua River in Missouri) have very low observed total-nitrogen concentrations, and the model did not fit these sites particularly well. These are the three sites with the lowest observed total nitrogen concentration values (fig. 2). These sites are not unique based on available national scale ancillary data other than the fact that they have low nitrogen concentrations for the given nonpoint source nitrogen inputs.

Nonagricultural Sites

Explanatory variables and regression coefficients in the CONDOR model for total-nitrogen concentrations downstream from nonagricultural areas are listed in table 3. In the simulation tests where two-thirds of the sites were randomly selected for model calibration, base flow index was nonsignificant eight times; other explanatory variables were always significant. For the model calibrated using data from all nonagricultural sites, the root-mean-square error is 0.25, the adjusted R-squared value is 0.62, and the bias-correction factor used for de-transforming estimated total-nitrogen concentrations is 1.20.

Models using manure, fertilizer, and atmospheric inputs as separate terms were not successful in explaining total-nitrogen concentrations in streams draining nonagricultural areas. Manure was typically a significant term but fertilizer inputs often were nonsignificant in the model development or subsequent model testing. Atmospheric deposition, which represents a much smaller nitrogen source, did not explain a sufficient amount of the variability in total-nitrogen concentrations. Combining the three nonpoint sources into one term resulted in a model that tested well, explained a significant proportion of the variability in total-nitrogen concentrations, and is useful in extrapolating results to areas where one of the three nonpoint-source terms may be dominant.

The form of the nonagricultural model is very simple in that it includes only source terms and flow variables. Population density acts as a surrogate for urban nutrient sources, such as wastewater-treatment plants and septic systems. Total-nitrogen concentrations tend to be higher at sites downstream from areas where population density or other nitrogen sources (manure accumulation, fertilizer application, and atmospheric deposition of nitrogen) are greater. Conversely, sites with

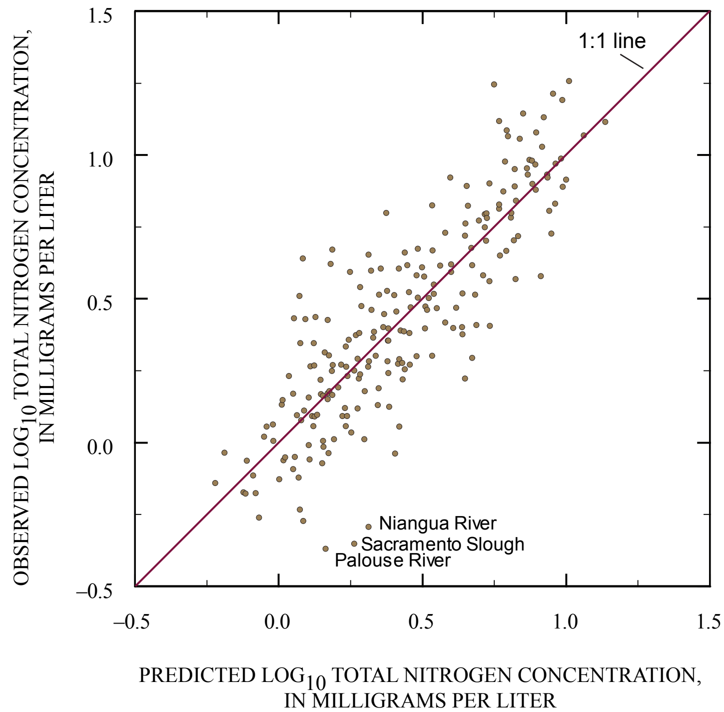


Figure 2. Comparison of observed and predicted flow-weighted mean annual total-nitrogen concentrations at National Water-Quality Assessment sampling sites downstream from agricultural areas.

Table 3. Explanatory variables in the model of $\log_{10}$ flow-weighted mean annual total-nitrogen concentrations downstream from nonagricultural areas.

[kg/km^2, kilograms per square kilometer; km^2, square kilometer]

Explanatory variable (transformation)	Regression coefficient	Standard error	p-value
Intercept	–0.00625	0.12281	0.9595
Nitrogen from manure plus fertilizer plus atmospheric inputs, kg/km^2 (fourth root)	0.09353	0.01661	<0.0001
Population density, per km^2 (fourth root)	0.09782	0.01241	<0.0001
Mean annual runoff 1951–80, inches ($\log_{10}$)	–0.55851	0.04065	<0.0001
Base flow index, percent of long-term mean streamflow	–0.00469	0.00124	<0.0001

greater runoff tend to have lower total-nitrogen concentrations, possibly due to dilution. Basins where a higher percentage of streamflow is generated during base-flow conditions also tend to have lower total-nitrogen concentrations. Although local conditions can vary, in general base flow might transport relatively less nitrogen than storm runoff. It should be noted that this model is applicable to areas with less than 25 percent agricultural land, where total-nitrogen concentrations in groundwater may not be elevated. Nolan and Hitt (2006) show that nitrate concentrations in shallow groundwater are directly correlated to agricultural land use (percent cropland and pastureland).

A comparison of observed and estimated total-nitrogen concentrations at nonagricultural sites is shown in figure 3. The regression fit for these sites is not as good as for the agricultural model, and less of the variation is explained. The site with the greatest observed nitrogen concentration in figure 3 is Las Vegas Wash near Las Vegas, Nevada. This site receives a large contribution of nutrients from a single point source (wastewater-treatment plant). Consistent point-source data

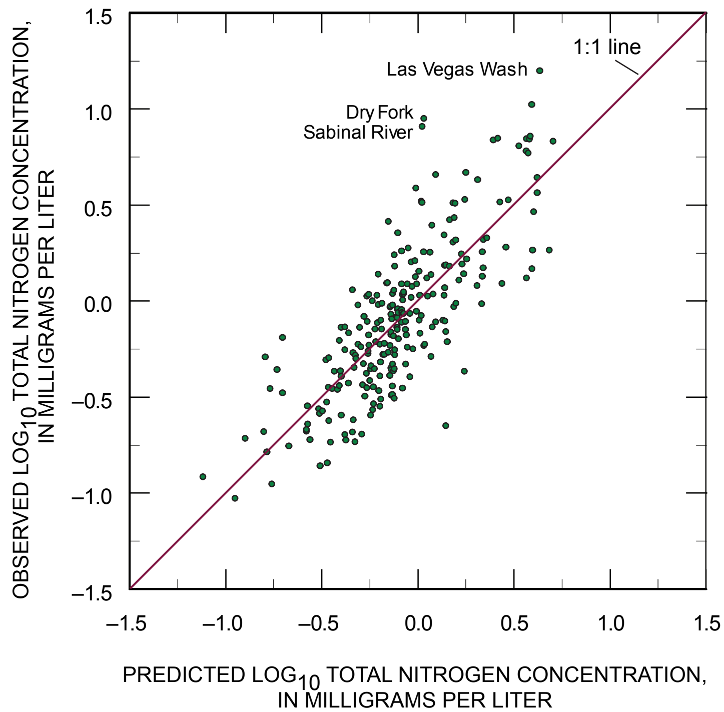


Figure 3. Comparison of observed and predicted flow-weighted mean annual total-nitrogen concentrations at National Water-Quality Assessment sampling sites downstream from nonagricultural areas.

were not available on a national scale, and none of the potential explanatory variables were useful in improving the estimation of total nitrogen at this site. Two other sites are seen above the scatter shown in figure 3 (Dry Fork of Roan Creek in Colorado and the Sabinal River in Texas). Streamflows can change very rapidly at these sites. During short-term high-flow events, large loads of nitrogen are exported from these basins. These high loads result in large flow-weighted total-nitrogen concentrations, and there is no term in the regression model to account for this type of event.

Analysis of Residuals

The geographic distributions of residuals from the total-nitrogen models for agricultural and nonagricultural sites are shown in figure 4. More sites fall into the plus or minus 0.25 log units range for the agricultural model compared to the nonagricultural model as a result of the difference in the overall error between the models. There is no clear geographic pattern evident in figure 4, sites with low to moderate residual values are found in all geographic regions, along with the occasional sites where the model did not accurately estimate total nitrogen.

An alternative way to investigate the regional pattern is shown in figure 5. Sites are classified by USEPA nutrient ecoregions (U.S. Environmental Protection Agency, 1998) and the distributions of residuals for sites in each ecoregion are shown by box plots. Due to the small number of sites in many nutrient ecoregions, the residuals for the agricultural and nonagricultural site models were grouped together. Eighty percent of the residuals for every nutrient ecoregion fall within plus or minus 0.5 log units. With the exception of ecoregions 4 and 14, where the regression models tend to underestimate, and ecoregion 7, where the models tend to overestimate, more than 50 percent of the residuals are within plus or minus 0.25 log units.

Basin area upstream from the sites used for model development ranged from 17 to 220,970 square kilometers (Mueller and Spahr, 2005). Sites were grouped into quartiles of upstream basin area (for example, sites downstream from the smallest 25 percent of basins are in quartile 1), and the distribution of regression model residuals for each quartile are shown in figure 6. The slightly better estimations for the agricultural model are evident by the smaller span of the box plots for the agricultural sites. The agricultural model estimates total nitrogen slightly more accurately for the largest basins (quartile 4). However, there are not large differences in the predictive capability of the models among different size basins.

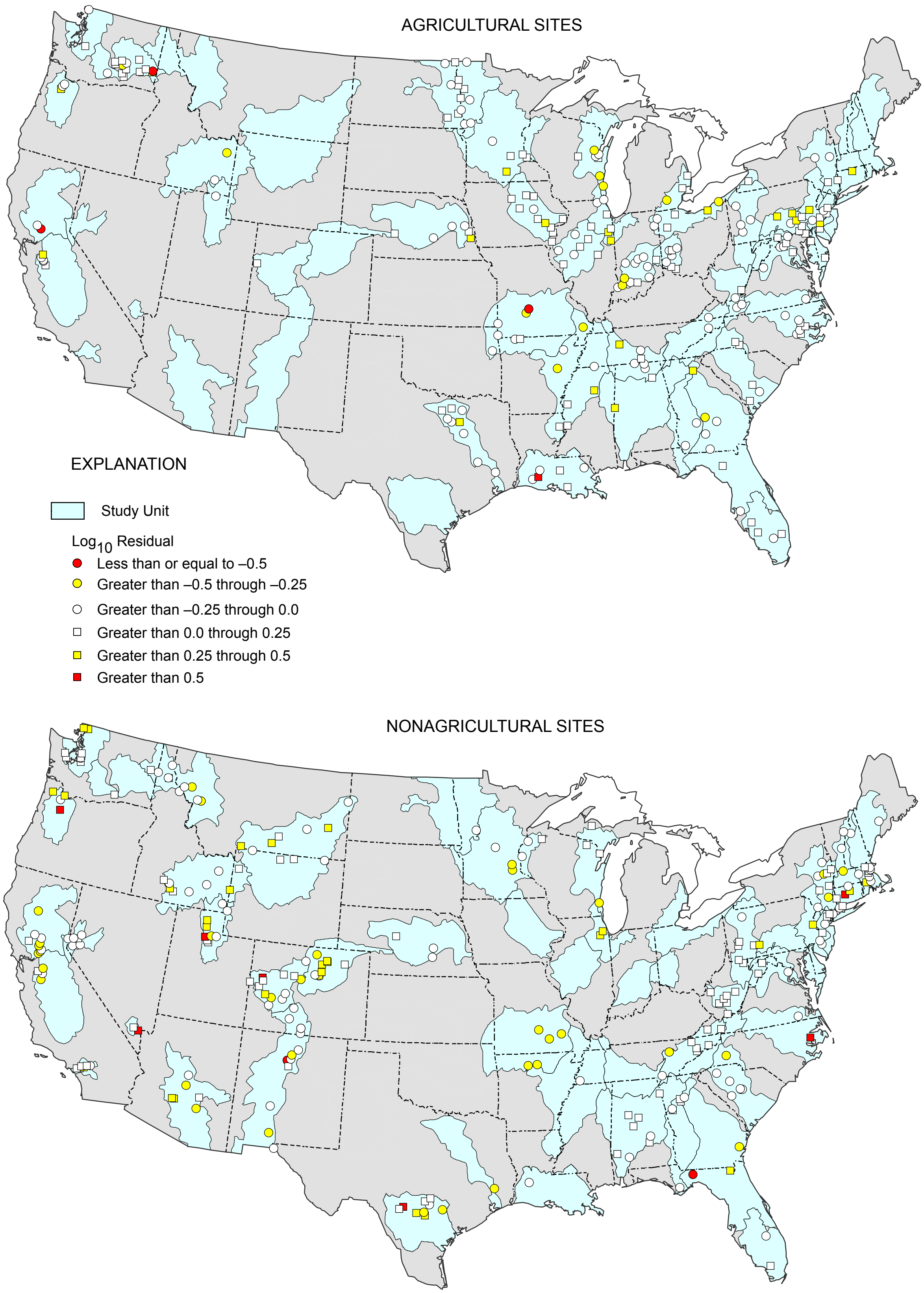


Figure 4. Geographic distribution of residuals for regression models of flow-weighted mean annual total-nitrogen concentrations at National Water-Quality Assessment sampling sites downstream from agricultural and nonagricultural areas.

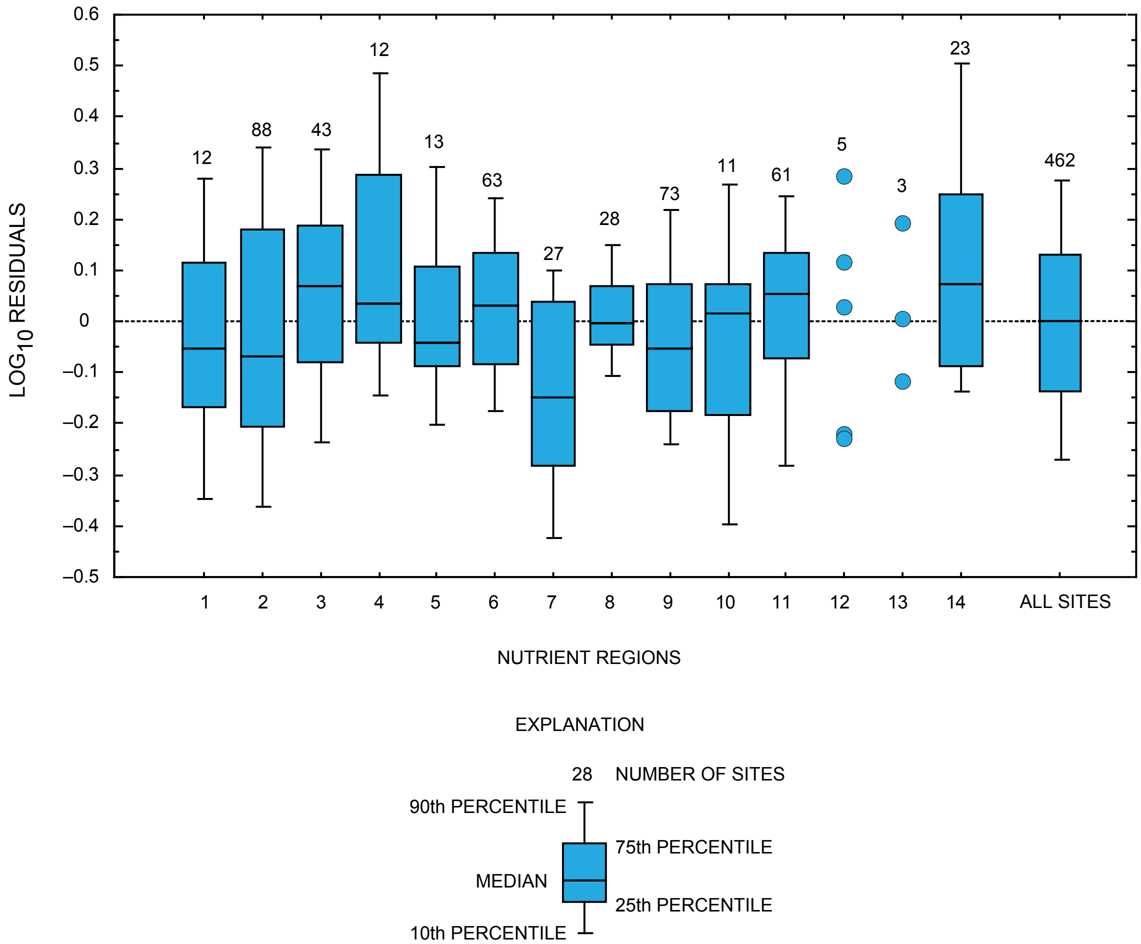


Figure 5. Distributions of residuals for regression models of flow-weighted mean annual total-nitrogen concentrations grouped by U.S. Environmental Protection Agency nutrient ecoregion (U.S. Environmental Protection Agency, 1998). Ecoregions 12 and 13 have insufficient sites to determine percentile for box plots.

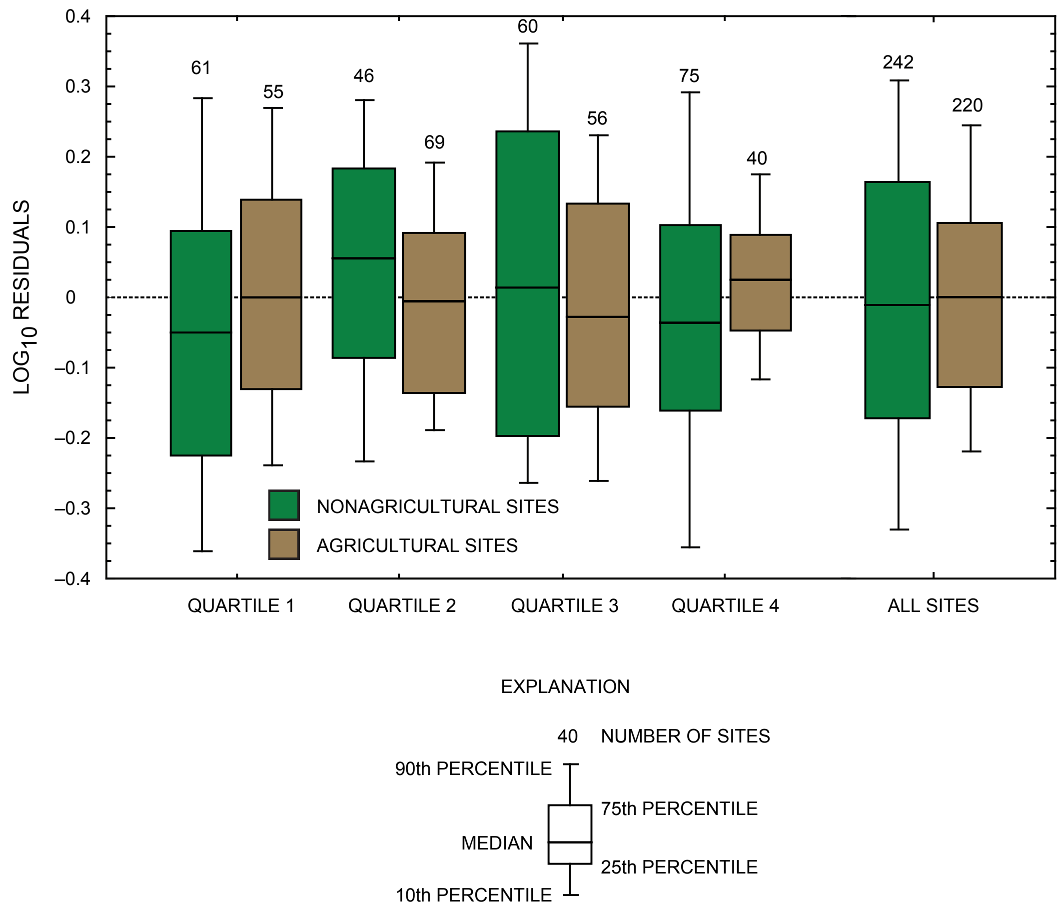


Figure 6. Distributions of residuals for regression models of flow-weighted mean annual total-nitrogen concentrations grouped by quartile of basin size.

Nationwide Prediction of Total-Nitrogen Concentrations

The calibrated regression models were used to predict flow weighted total-nitrogen concentrations in streams throughout the conterminous United States (fig. 7). The values represent mean annual flow weighted concentrations for the Cycle I period (1992–2001) of the NAWQA program. Explanatory variables in the models were evaluated based on geographic-information-system data for basins upstream from each node in the RF1 file. Each basin was classified as either agricultural or nonagricultural so the appropriate regression model could be applied. Model applications produced a predicted mean-annual concentration of total nitrogen at each RF1 node. The concentration predicted at a node was assumed to occur in the entire reach upstream to the next node.

Predicted total-nitrogen concentrations for about 37 percent of the stream miles throughout the Nation were less than 1 mg/L, and predictions were less than 5 mg/L for another 54 percent of stream miles. Predicted concentrations less than 0.5 mg/L occurred primarily in mountainous regions: the Rocky Mountain, Sierra Nevada, Cascade, and Coastal ranges in the West; and the Appalachian and Adirondack ranges in the East. Low concentrations also occurred in northern areas around the Great Lakes and Maine. Concentrations greater than 5 mg/L were predicted within a broad area extending from Ohio to eastern Nebraska. Agricultural application of fertilizer and manure is high throughout this area, and atmospheric deposition of nitrogen is high in the eastern part. Predicted concentrations also are high around urban areas in California, Arizona, Colorado, and Texas. The lack of a national data set for point sources of nutrients and the subsequent unavailability of a true point-source term in the regression equations results in an under representation of urban areas in the national prediction (the residuals from the calibration of the nonagricultural sites model showed a positive bias for urban basins).

Mean annual concentrations of total nitrogen exceeding the USEPA drinking-water standard for nitrate (10 mg/L; U.S. Environmental Protection Agency, 2003) were predicted for 1 percent of stream miles. Because nitrate is a component of the total nitrogen concentration value, the nitrate concentration in water always will be less than the concentration of total nitrogen. Figure 7, therefore, indicates that widespread exceedence of the nitrate drinking water standard in streams is unlikely.

The predicted concentrations have an associated uncertainty as a result of variability in the observations that is unexplained by the regression models. This uncertainty can be used to estimate the probability that a predicted concentration exceeds a reference value or any other specified concentration (see equation 1). If the predicted value is equal to a reference concentration there is a 50 percent probability that the actual flow-weighted annual mean concentration in the stream reach is less than the reference value and a 50 percent probability that the actual concentration is greater than the reference concentration. To illustrate the uncertainty of the regression models

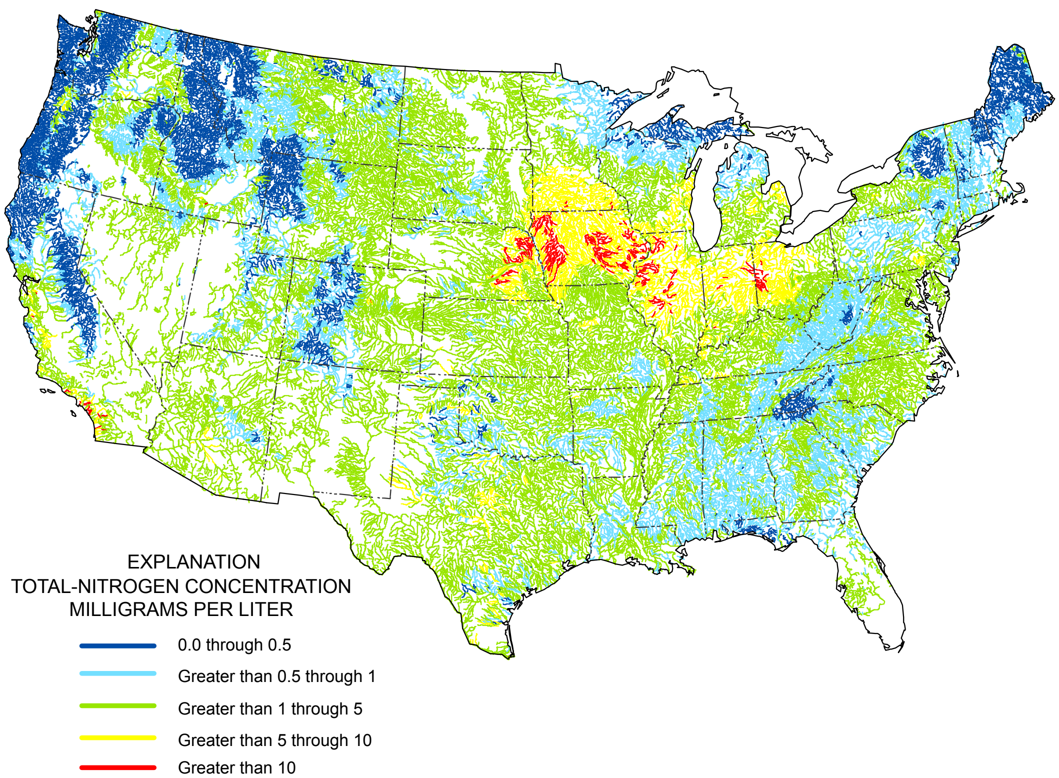


Figure 7. Predicted flow-weighted mean-annual concentrations of total nitrogen for streams in the U.S. Environmental Protection Agency River Reach File, 1992–2001.

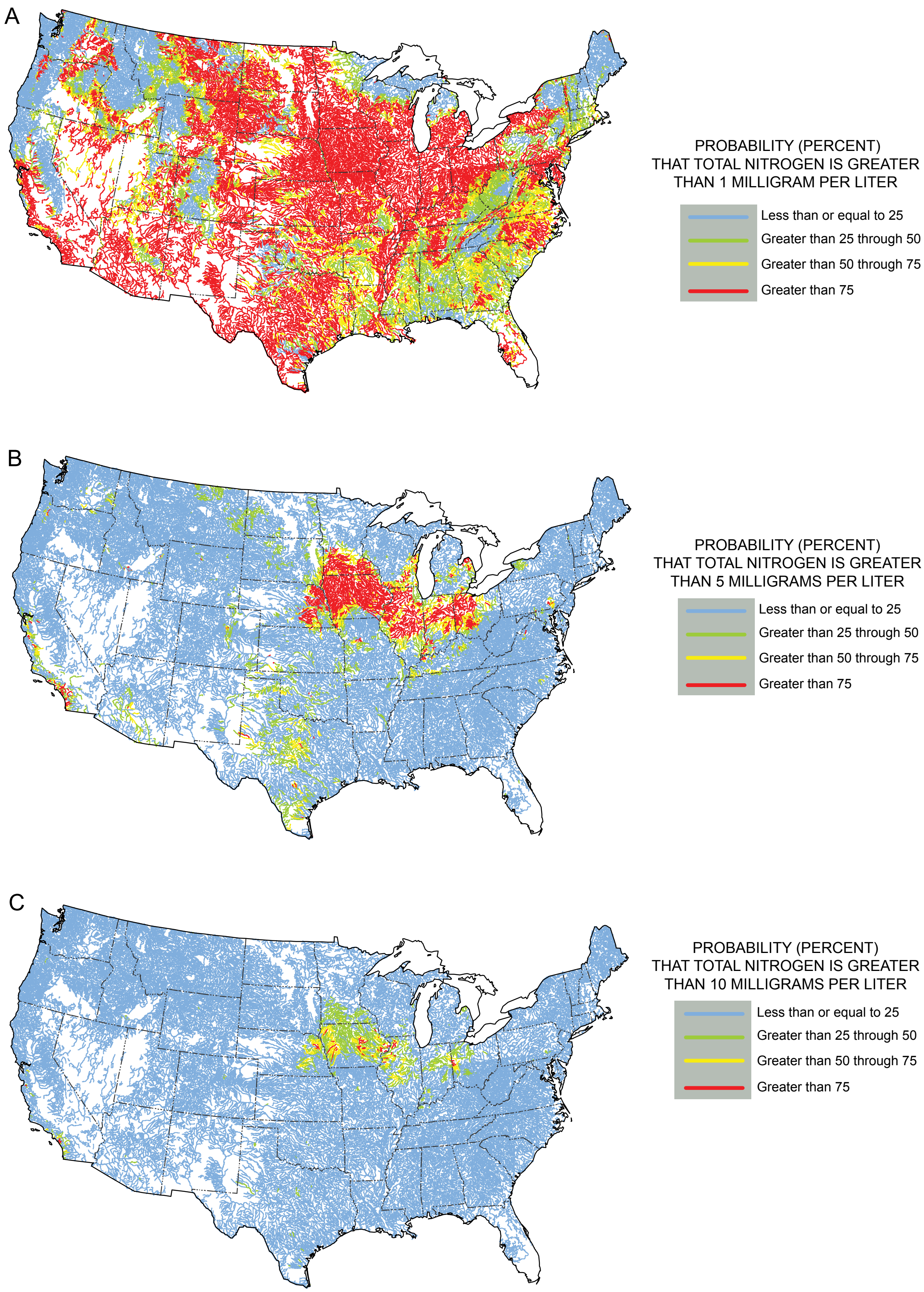


Figure 8. Probabilities that flow-weighted mean-annual concentrations of total nitrogen in streams are greater than 1, 5, and 10 milligrams per liter.

in the predicted concentrations, probabilities of concentrations being greater than 1, 5, and 10 mg/L are shown in figure 8.

The reaches colored red in figure 8A show the extensive areas where the probability is at least 75 percent that total-nitrogen concentration exceeds 1 mg/L. Figure 8B shows far fewer areas where total-nitrogen concentration in stream water is likely to exceed 5 mg/L. The probability that the concentration exceeds 5 mg/L is at least 75 percent in many agricultural areas of the Midwest and near some urban areas in the West. In most other areas, the probability of total-nitrogen concentrations exceeding 5 mg/L is no more than 25 percent. The probability of total-nitrogen concentrations exceeding 10 mg/L is low everywhere except for isolated areas in the Midwest and California (fig. 8C).

The probabilities that mean-annual concentrations of total nitrogen in streams exceed USEPA regional nutrient criteria (U.S. Environmental Protection Agency, 2002a) are shown in figure 9. Each RF1 stream reach was associated with the appropriate USEPA nutrient ecoregion. The predicted concentration in the reach was compared to the regional criterion and the probability that the estimated concentration exceeded the criterion was computed, based on the regression-model prediction error. In the vast majority of streams, the total-nitrogen concentration is likely to exceed the criteria (probability at

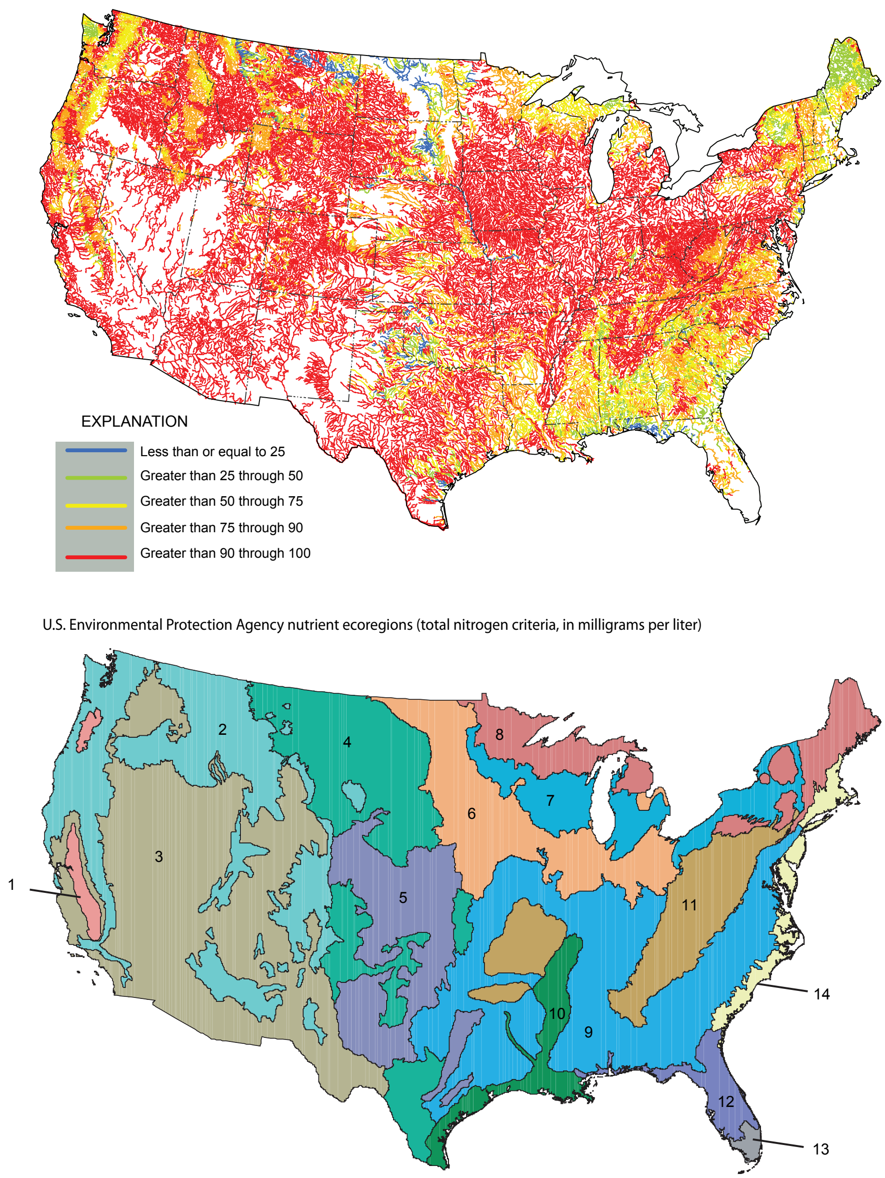


Figure 9. Probabilities that flow-weighted mean-annual concentrations of total nitrogen in streams exceed the U.S. Environmental Protection Agency nutrient-ecoregion criteria.

least 50 percent). These exceedence probabilities are determined for mean annual concentrations; the probability of a shorter term exceedence is even greater. For many streams the probability of exceeding the regional criteria is more than 90 percent. Streams where the probability of exceeding the criteria is low are located primarily in the northern-most States and the Southeast, but in most of these streams there is still a 25 to 50 percent probability that these concentrations could exceed the regional criteria.

The percent of river miles likely to exceed the total nitrogen criteria in each nutrient ecoregion is shown in figure 10. Results are not available for nutrient region 13, Southern Florida Coastal Plain, because nutrient criteria are unavailable for this area. In all other nutrient regions, the probability of exceeding the criteria is at least 50 percent for more than half of the stream miles. With the exception of nutrient regions 8, 12, and 14, there is a 90 percent probability that the total nitrogen criteria will be exceeded in more than 48 percent of the stream miles. In nutrient regions 3 and 7 the nitrogen criterion is likely to be exceeded in almost 80 percent of streams with a probability of at least 90 percent.

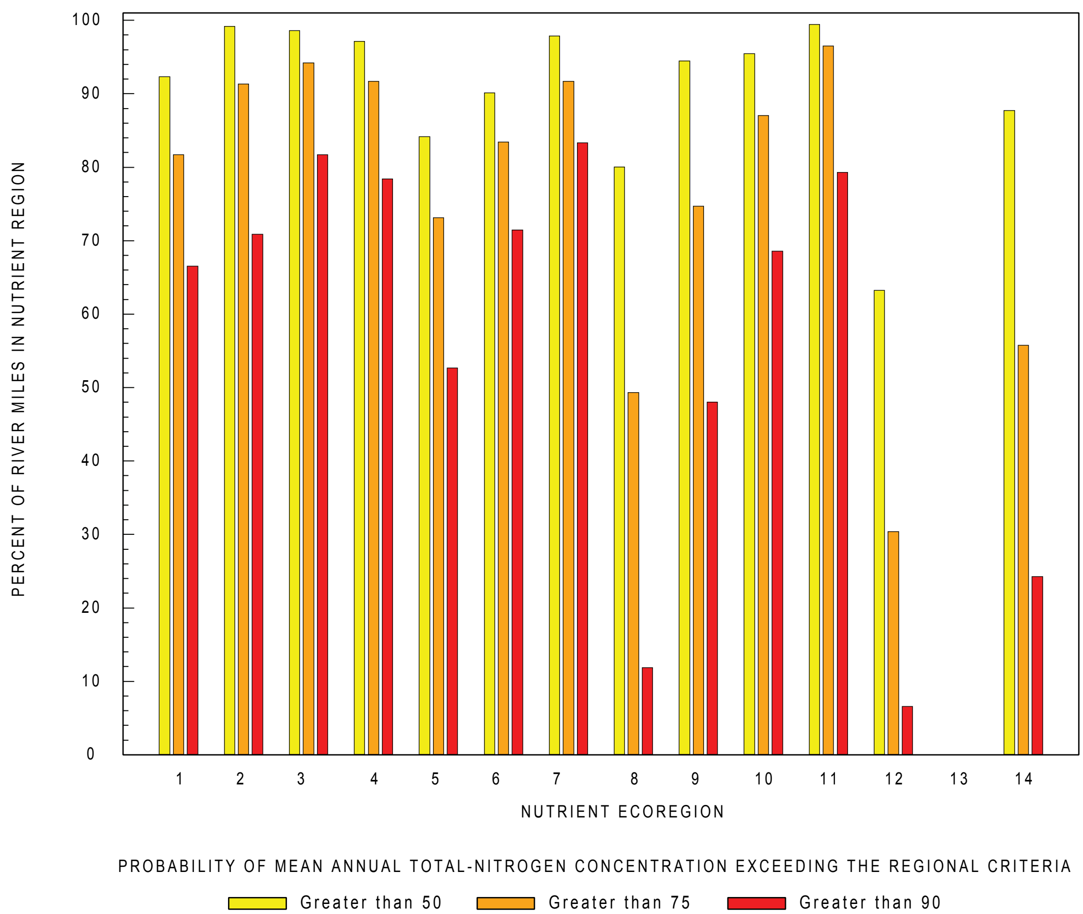


Figure 10. Percentage of stream miles in each U.S. Environmental Protection Agency nutrient-ecoregion within which the probabilities of exceeding the regional total-nitrogen criteria are 50, 75, and 90 percent (criteria not available for ecoregion 13).

The total-nitrogen models presented herein are statistical models based on correlations of nonpoint source inputs, basin characteristics, and land use. These models provide reasonable estimates of total-nitrogen concentrations in streams that do not have an abundance of measured data. The models can be used to provide broad regional estimates of how nitrogen concentrations may change with reductions in nonpoint sources of nitrogen. To investigate the regional response of stream nitrogen concentrations to reductions in nonpoint sources of nitrogen, fertilizer inputs were reduced by specified amounts and the percent of river miles with a 90-percent probability of exceeding the nutrient criteria was computed (fig. 11). It should be noted that the models are not being used to specifically infer what may change in any particular stream but, rather, what may occur on a broad regional scale. The results suggest a very small reduction in the number of stream miles with a 90-percent probability of exceeding the criteria, even when fertilizer inputs are reduced by 25 percent.

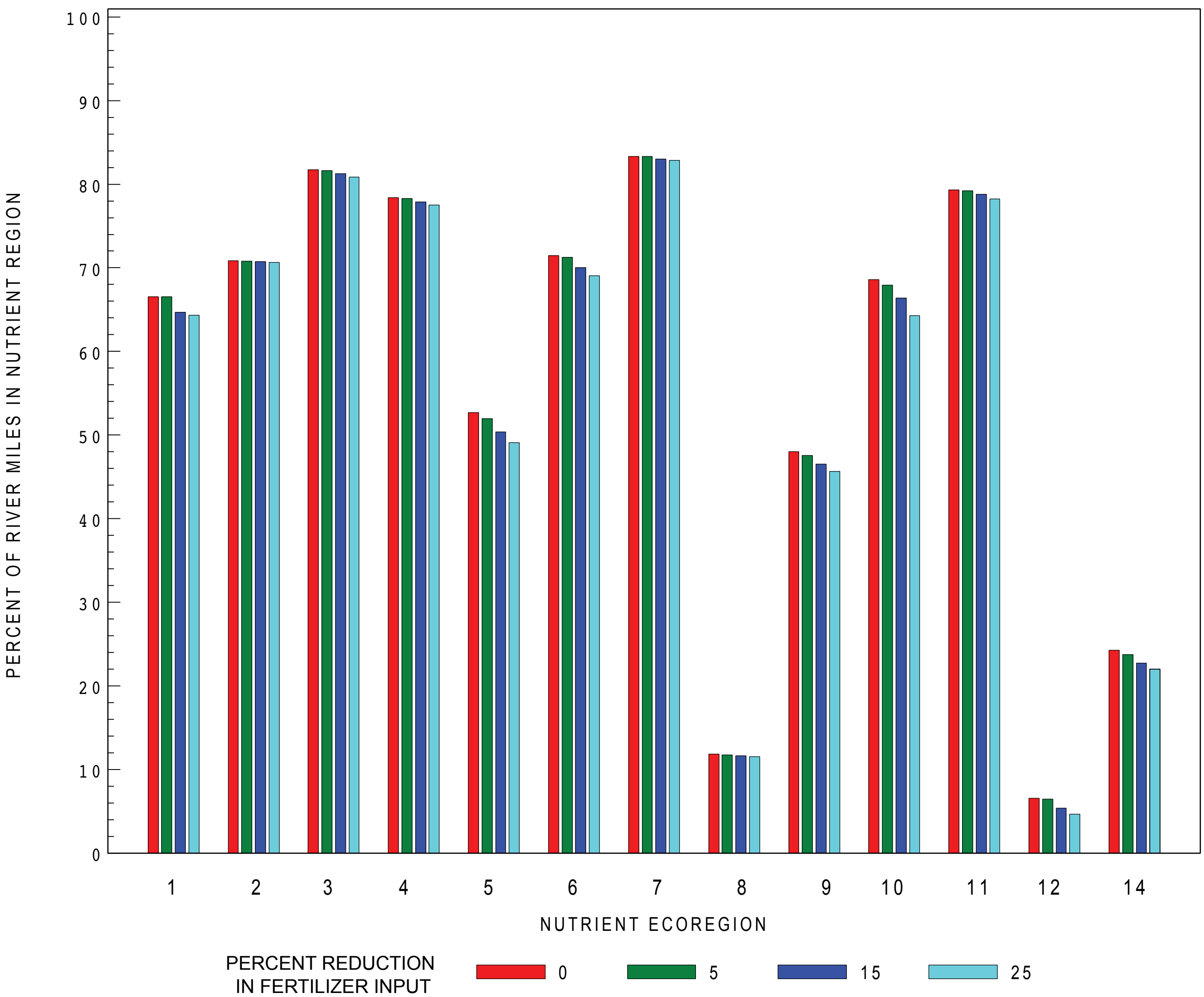


Figure 11. Percentage of stream miles in each U.S. Environmental Protection Agency nutrient-ecoregion with a 90-percent probability of exceeding the regional total-nitrogen criteria with various reductions in fertilizer inputs.

Development of Models for Total Phosphorus

Attempts to develop national regression models to estimate total phosphorus concentrations had limited success. Unlike the total nitrogen models, division of sites into agricultural and nonagricultural groups did not improve the explanatory capability of the equations. The group of explanatory variables that yielded the lowest model error is listed in table 4. For this model, calibrated using data from 440 NAWQA sites, the root-mean-square error is 0.36 and the adjusted R-squared value is 0.52.

Significant source terms include nonpoint-source inputs from manure and population density, which is acting as a surrogate for point-source inputs as well as other nonpoint urban sources. Nonpoint-source input of phosphorus from fertilizer was not a significant explanatory variable at the national scale, perhaps because phosphorus tends to adsorb to soil particles rather than dissolve in water contributing to streamflow or groundwater. Soil erosion and soil phosphorus also were not significant explanatory variables; however, range land use was significant and might be functioning as a surrogate for soil erosion. Other studies (Peterson and others, 2004; Mueller and Spahr, 2006) have shown a correlation between range land use and phosphorus concentrations, probably as a result of mineral and biological sources of phosphorus on range lands and transport of particulate phosphorus to streams along with sediment. In areas without rangeland, lower phosphorus concentrations are associated with areas of less development as indicated by the negative coefficient for the forested land-use variable. Percent sand also has a negative coefficient. More sand in soil can increase the infiltration rate, leading to less sediment and associated phosphorus transport to streams. Similar to the model for total nitrogen at nonagricultural sites, base-flow index was negatively correlated with total phosphorus. At sites with a low percentage of baseflow, there is a greater likelihood of runoff events, which transport sediment and associated particulate phosphorus to the stream.

Observed compared to estimated mean-annual concentrations of total phosphorus and the distribution of the regression model residuals are shown in figure 12. The scatter about the unity line for total phosphorus is greater than the scatter for the total-nitrogen models (figs. 2 and 3). The residuals are well distributed about zero, but the range indicates a large amount of variability in total-phosphorus concentrations is not explained by the model. The degree of unexplained variability precludes the use of this model for nationwide extrapolation because the predicted values would have large uncertainty and the resulting geographic patterns of phosphorous concentrations could be misleading.

Table 4. Explanatory variables in the model of $\log_{10}$ flow-weighted mean annual total-phosphorus concentrations.

[kg/km^2, kilograms per square kilometer; km^2, square kilometer]

Explanatory variable (transformation)	Regression coefficient	Standard error	p-value
Intercept	–0.31618	0.16884	0.0618
Manure phosphorus input, kg/km^2 (fourth root)	0.08150	0.01431	<.0001
Population density, per km^2 (fourth root)	0.08482	0.01547	<.0001
Range land use, percent of basin area	0.00560	0.00096310	<.0001
Forest land use, percent of basin area	–0.00731	0.00075089	<.0001
Sand in soil, percent (fourth root)	–0.23044	0.06481	0.0004
Base flow index, percent of long-term mean streamflow	–0.00701	0.00132	<.0001

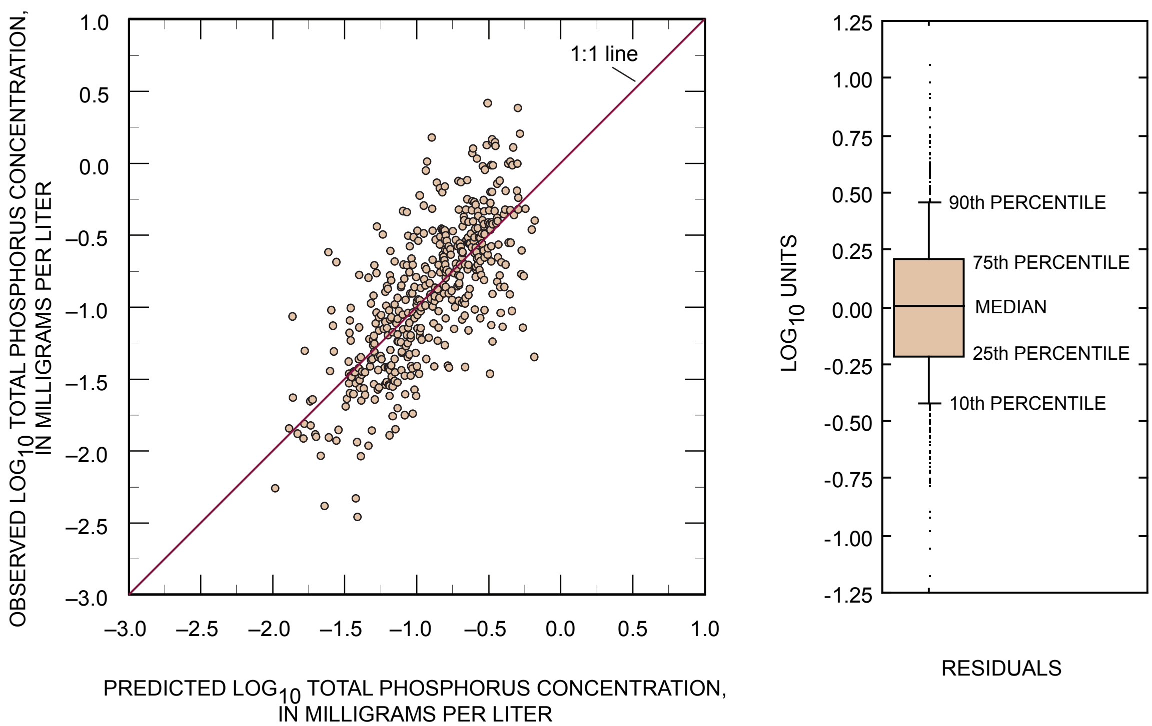


Figure 12. Comparison of observed and predicted total-phosphorus concentrations and the distribution of residuals from the regression model.

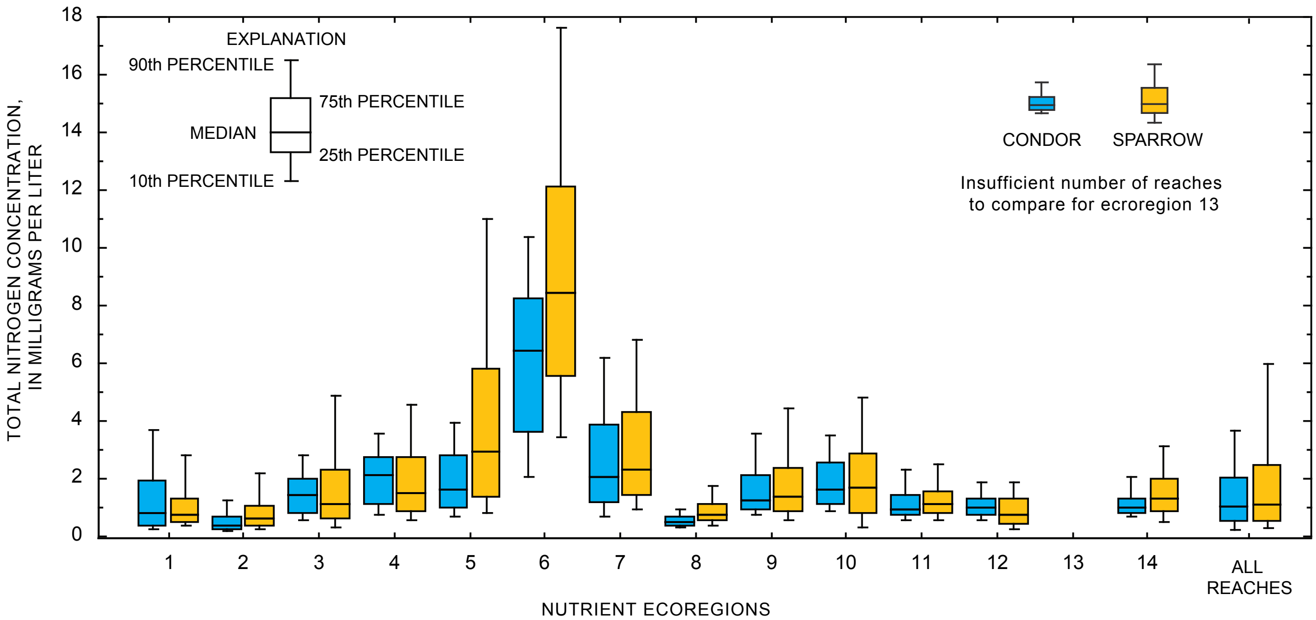
Comparison to Results from other Models

Smith and others (1997) calibrated statistically based models of total-nitrogen and total-phosphorus loads for streams in the conterminous United States using data from the USGS National Stream-Quality Accounting Network (NASQAN). Collectively, the models are referred to as a spatially referenced regression on watershed attributes (SPARROW) model of nutrient transport. The SPARROW model has since been used for numerous purposes, including estimation of natural nutrient background concentrations and comparison to USEPA criteria (Smith and others, 2003). Predictions of total-nitrogen load from a revised national SPARROW model (Alexander and others, 2008) were converted to concentrations using long-term mean runoff (R.B. Alexander, U.S. Geological Survey, written commun., 2008). These SPARROW -based concentrations were compared by river reach to predicted total-nitrogen concentrations from the CONDOR model.

Nutrient ecoregions were chosen as a geographic framework to summarize the comparison of CONDOR and SPARROW model results (fig. 13). The geographic distribution of areas with high and low concentrations is similar for the predicted concentrations from the CONDOR and SPARROW models. Regions with high predicted concentrations using CONDOR (ecoregions 5, 6, and 7) also show high concentrations for SPARROW. Both models also predict low concentrations in the Western Forested Mountains and Nutrient-Poor, Largely Glaciated Upper Midwest and Northeast (ecoregions 2 and 8). SPARROW results for these regions are also among the lowest for the country.

Although the regional patterns are similar between the two models, SPARROW results are typically greater than CONDOR predictied concentrations in many areas, and the differences are greater in areas of higher concentrations. The largest differences are found in the western part of the Corn Belt (in ecoregion 6) and the eastern part of the South Central Cultivated Great Plains (ecoregion 5). Large relative differences were also found in areas of low concentrations. Predicted concentrations using CONDOR in ecoregions 2 and 8 (Western Forested Mountains and Nutrient-Poor, Largely Glaciated Upper Midwest and Northeast) are less than concentrations derived from SPARROW.

There are differences between the model structures of CONDOR and SPARROW that could cause differences in the estimated nitrogen concentrations. The SPARROW model uses nonlinear regression to solve a mass balance of nutrient loads as a function of spatially referenced nutrient inputs, land-use delivery proportions, and instream decay characteristics along the river reach. CONDOR, in contrast, uses linear regression to relate nutrient concentrations measured at stream sites to the drainage basin characteristics (such as nutrient inputs, soil and hydrologic variables) of those sites. SPARROW uses data from the NASQAN network for the years 1975–2000 with a base year of 1992 to derive the loads used for model calibration, whereas CONDOR is based on concentration, streamflow, and load data collected during the 3- or 4- year high-intensity phase for sites within each of the Cycle I NAWQA study units during 1992–2001.



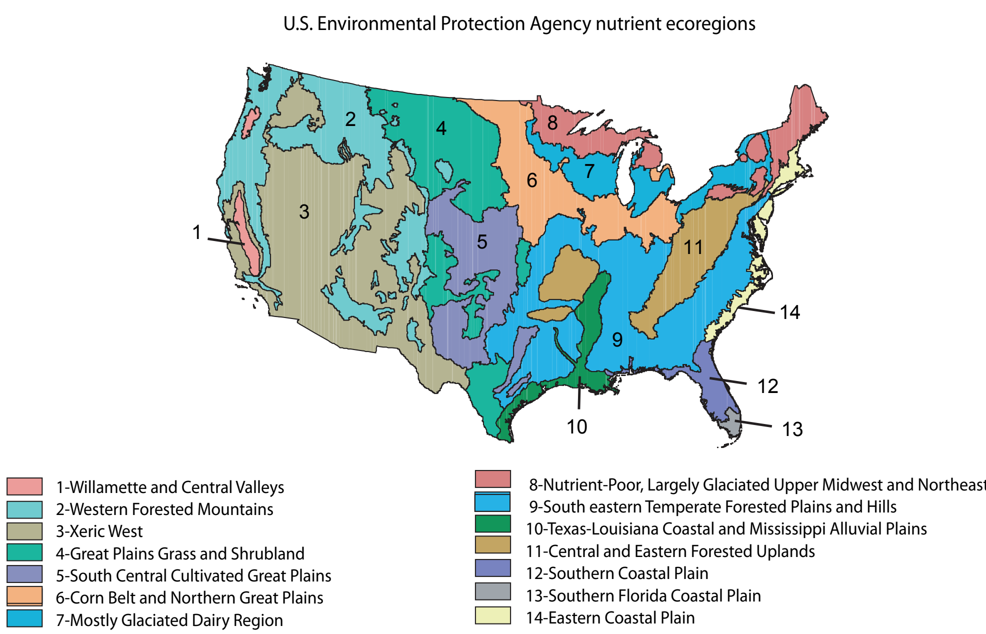


Figure 13. Comparison of prediction total-nitrogen concentrations from this report and the SPARROW model (Alexander and others, 2008).

Summary and Conclusions

Nutrients (compounds containing nitrogen or phosphorus) are one of the primary synthesis topics of the USGS NAWQA program. During the Cycle I (1992–2001) of the NAWQA program, water samples were collected at a national network of stream sites to quantify the occurrence and distribution of nitrogen and phosphorus. Extrapolation of these data to unsampled areas was accomplished through the use of statistical regression models relating flow-weighted mean annual concentrations to nonpoint sources, basin characteristics, and hydrologic conditions. Explanatory variables are limited to data that are nationally comprehensive. Higher resolution localized data were not applicable to this analysis due to the limited spatial extent. For example, point-source loading of nitrogen and phosphorus is unavailable in a nationally consistent form and thus could not be used to assist in the explanation of observed nutrient concentrations at a national scale.

The concentrations of nutrients derived from regression (CONDOR) models were used to predict annual flow-weighted mean concentrations of total nitrogen, representative of the Cycle I period of the NAWQA program, for drainage basins associated with the USEPA River Reach File. Manure, fertilizer, and atmospheric inputs of nitrogen, percent sand in soil, subsurface drainage, overland flow, mean annual precipitation, and percent undeveloped area were significant variables in the agricultural basin total nitrogen model. Significant explanatory variables in the nonagricultural total nitrogen model were total nonpoint-source nitrogen input (sum of nitrogen from manure, fertilizer, and atmospheric deposition), population density, mean annual runoff, and percent base flow. The majority of stream miles in the Nation have predicted concentrations less than 5 mg/L. Concentrations greater than 5 mg/L were predicted for a broad area extending from Ohio to eastern Nebraska, areas spatially associated with greater application of fertilizer and manure. Predicted concentrations also are high around urban areas in California, Arizona, Colorado, and Texas. The lack of a national data set for point sources of nutrients and the subsequent unavailability of a true point-source term in the regression equations results in an under representation of urban areas in the national prediction. Mean annual concentrations of total nitrogen exceeding the USEPA drinking water standard for nitrate (10 mg/L) were predicted for less than 1 percent of stream miles. Because nitrate is a nitrogen compound, its concentration in water can be no more than the concentration of total nitrogen. Thus, widespread exceedence of the drinking water standard in streams is unlikely in most of the Nation.

Predicted concentrations have an associated uncertainty as a result of variability unexplained by the regression models. Probabilities that concentrations will exceed a given reference value can be determined by incorporating model uncertainty. The probabilities that mean-annual concentrations of total nitrogen in streams exceed USEPA regional nutrient criteria were determined for reaches in the RF1 River Reach File. In all nutrient regions where criteria have been established, there is at least a 50 percent probability of exceeding the criteria in more than half of the stream miles.

Dividing calibration sites into agricultural and nonagricultural groups did not improve the explanatory capability for total phosphorus models. The group of explanatory variables that yielded the lowest model error for mean annual total phosphorus concentrations includes phosphorus input from manure, population density, amounts of range land and forest land, percent sand in soil, and percent base flow. The national regression model to estimate total-phosphorus concentration had limited success. Available national-scale explanatory variables did not adequately explain annual total-phosphorus concentrations. The unexplained variability precluded the use of the model for nationwide prediction because the predicted values would have an unreasonably large uncertainty.

Acknowledgments

This report relies on water samples that were collected over a 10-year period by hydrologists and hydrologic technicians in 51 NAWQA study units and analyzed by technicians at the USGS National Water Quality Laboratory in Denver, Colorado. Without all their conscientious efforts, the data interpretation in this report would not have been possible.

Richard Moore, Wesley Stone, and Betty Palcsak of the U.S. Geological Survey provided valuable comments on the report. The authors also thank Chelsea Carr for editorial review and Mary Berger for manuscript preparation, layout, and illustration assistance.

References Cited

Alexander, R.B., Smith, R.A., Schwarz, G.E., Boyer, E.W., Nolan, J. V., and Brakebill, J.W., 2008, Differences in phosphorus and nitrogen delivery to the Gulf of Mexico from the Mississippi River Basin: Environmental Science and Technology, v. 2, no. 3, p. 822–830.

Daly, C., Taylor, G.H., Gibson, W. P., Parzybok, T.W., Johnson, G. L., and Pasteris, P., 2001, High-quality spatial climate data sets for the United States and beyond: Transactions of the American Society of Agricultural Engineers, v. 43, p. 1957–1962.

Duan, N., 1983, Smearing estimate-A nonparametric retransformation method: Journal of the American Statistical Association, v. 78, p. 605–610.

Gebert, W.A., Graczyk, D.J., and Krug, W.R., 1987, Average annual runoff in the United States: U.S. Geological Survey Hydrologic Atlas 710.

Gilliom, R.J., Alley, W.M., and Gurtz, M.E., 1995, Design of the National Water-Quality Assessment Program—Occurrence and distribution of water-quality conditions: U.S. Geological Survey Circular 1112, 33 p.

Gilliom, R.J., Barbash, J.E., Crawford, C.G., Hamilton, P.A., Martin, J.D., Nakagaki, Naomi, Nowell, L.H., Scott, J.C., Stackelberg, P.E., Thelin, G.P., Wolock, D.M., 2006, Pesticides in the Nation’s streams and ground water, 1992–2001: U.S. Geological Survey Circular 1291, 172 p.

Helsel, D.R., and Hirsch, R.M, 1993, Statistical methods in water resources: New York, Elsevier Science Publishers B.V., 529 p.

Mueller, D.K., and Spahr, N.E., 2005, Water-quality, streamflow, and ancillary data for nutrients in streams and rivers across the Nation, 1992–2001: U.S. Geological Survey Data Series 152, accessed December 2005, at: http://pubs.usgs.gov/ds/2005/152/.

Mueller, D.K., and Spahr, N.E., 2006, Nutrients in streams and rivers across the nation—1992–2001: U.S. Geological Survey Scientific Investigations Report 2006–5107, 44 p.

Nakagaki, Naomi, and Wolock, D.M., 2005, Estimation of agricultural pesticide use in drainage basins using land cover maps and county pesticide data: U.S. Geological Survey Open-File Report 2005–1188, 46 p.

Natural Resources Conservation Service, 1994, State soil geographic (STATSG0) data base—Data use information: U.S. Department of Agriculture, National Soil Survey Center, Miscellaneous Publication Number 1492, 110 p.

Natural Resources Conservation Service, 1995, 1992 National resources inventory: U.S. Department of Agricultural, Natural Resources Conservation Service.

Nolan, B.T. and Hitt, K.J., 2006, Vulnerability of shallow groundwater and drinking-water wells to nitrate in the United States: Environmental Science and Technology, v. 40, no. 25, p. 7,834–7,840.

Nolan, J.V., Brakebill, J.W., Alexander, R.B., and Schwarz, G.E., 2003, Enhanced River Reach File 2.0, Version 2.0, November 10, 2003: U.S. Geological Survey Open-File Report 2002–40.

Nowell, L.H., Crawford, C.G., Nakagaki, N., Thelin, G.P., and Wolock, D.M., 2006, Regression model for explaining and predicting concentrations of dieldrin in whole fish from U.S. streams: U.S. Geological Survey Scientific Investigations Report 2006–5020, 30 p.

Peterson, D.A., Miller, K.A., Bartos, T.T., Clark, M.L., Porter, S.D., and Quinn, T.L., 2004, Water quality in the Yellowstone River Basin, Wyoming, Montana, and North Dakota, 1999–2001; U.S. Geological Survey Circular 1234, 34 p.

Ruddy, B.C., Lorenz, D.L., and Mueller, D.K., 2006, County-level estimates of nutrient inputs from fertilizer, manure, and atmospheric-deposition sources in the conterminous United States, 1982–2001: U.S. Geological Survey Scientific Investigations Report 2006–5012, 17 p.

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.

SAS Institute Inc., 1999, SAS procedures guide, version 8: Cary, N.C., SAS Institute Inc., 1643 p.

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.

Smith, R.A., Alexander, R.B., and Schwarz , G.E., 2003, Natural background concentrations of nutrients in streams and rivers of the conterminous United States: Environmental Science and Technology, v. 37, no. 14, p. 3,039–3,047.

U.S. Bureau of the Census, 1991a, Census of population and housing, 1990: Public Law (P.L.) 94–171 data on CD–ROM: U.S. Bureau of the Census, Washington D.C., digital tabular data.

U.S. Bureau of the Census, 1991b, TIGER/Line Census Files, 1990: U.S. Bureau of the Census, Washington D.C., digital tabular data.

U.S. Environmental Protection Agency, 1998, National strategy for the development of regional nutrient criteria: U.S. Environmental Protection Agency Office of Water Fact Sheet EPA–822–F–98–002, accessed January 2005, at: http://epa.gov/waterscience/criteria/nutrient/nutsi.html.

U.S. Environmental Protection Agency, 2002a, Summary table for the nutrient criteria documents: U.S. Environmental Protection Agency Office of Water, accessed January 2005, at: http://www.epa.gov/waterscience/criteria/nutrient/ecoregions/sumtable.pdf.

U.S. Environmental Protection Agency, 2002b, National water quality inventory—2000 report: U.S. Environmental Protection Agency Office of Water Report EPA–841–R–02–001, 207 p., accessed January 2007, at http://www.epa.gov/305b/2000report/.

U.S. Environmental Protection Agency, 2003, National primary drinking water standards: U.S. Environmental Protection Agency Office of Water Report EPA 816–F–03–016, accessed May 2006, at: http://www.epa.gov/safewater/mcl.html#mcls.

U.S. Environmental Protection Agency, 2007, National water quality inventory—Report to Congress, 2002 reporting cycle: U.S. Environmental Protection Agency Office of Water Report EPA–841–R–07–001, 39 p., accessed November 2007, at http://www.epa.gov/305b/2002report/.

Wolock, D.M., 2003a, Base-flow index grid for the conterminous United States: U.S. Geological Survey Open-File Report 2003–263, accessed March 2009, at http://pubs.er.usgs.gov/usgspubs/ofr/ofr03263.

Wolock, D.M., 2003b, Saturation overland flow estimated by TOPMODEL for the conterminous United States: U.S. Geological Survey Open-File Report 2003–264, accessed March 2009, at http://pubs.er.usgs.gov/usgspubs/ofr/ofr03264.

Publishing support provided by:
Denver Publishing Service Center

For more information concerning this publication, contact:
Director, USGS Colorado Water Science Center
Box 25046, Mail Stop 415
Denver, CO 80225
(303) 236-4882

Or visit the Colorado Water Science Center Web site at:
http://co.water.usgs.gov/

USGS

Spahr and others—**Development and Application of Regression Models for Nutrients in U.S. Streams**—Scientific Investigations Report 2009–5199

ISBN 978-141132595-1

9 781411 325951

Printed on recycled paper