science for a changing world

Evaluation and Use of U.S. Environmental Protection Agency Clean Watersheds Needs Survey Data to Quantify Nutrient Loads to Surface Water, 1978–2012

Scientific Investigations Report 2017–5115

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

Evaluation and Use of U.S. Environmental Protection Agency Clean Watersheds Needs Survey Data to Quantify Nutrient Loads to Surface Water, 1978–2012

By Tamara Ivahnenko

Scientific Investigations Report 2017–5115

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

U.S. Department of the Interior
RYAN K. ZINKE, Secretary

U.S. Geological Survey
William H. Werkheiser, Acting Director

U.S. Geological Survey, Reston, Virginia: 2017

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

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

Suggested citation:
Ivahnenko, Tamara, 2017, Evaluation and use of U.S. Environmental Protection Agency Clean Watersheds Needs Survey data to quantify nutrient loads to surface water, 1978–2012: U.S. Geological Survey Scientific Investigations Report 2017–5115, 11 p., https://doi.org/10.3133/sir20175115.

ISSN 2328-0328 (online)

Contents

Figures

Tables

Conversion Factors

U.S. customary units to International System of Units

Multiply	By	To obtain
	Volume	
gallon (gal)	3.785	liter (L)
million gallons (Mgal)	3,785	cubic meter (m^3)

International System of Units to U.S. customary units

Multiply	By	To obtain
	Mass	
kilogram (kg)	2.205	pound avoirdupois (lb)
liter (L)	0.2642	gallon (gal)

Abbreviations

CWNS	Clean Watersheds Needs Survey
EPA	U.S. Environmental Protection Agency
GIS	geographic information system
ICIS	Integrated Compliance Information System
MGD	million gallons per day
NAWQA	National Water Quality Assessment
NPDES	National Pollutant Discharge Elimination System
NPS	nonpoint source
PCS	Permit Compliance System
SIC	Standard Industrial Classification
SPARROW	Spatially Referenced Regressions on Watershed attributes
TN	total nitrogen
TP	total phosphorous
TPC	typical pollutant concentration
USGS	U.S. Geological Survey
WWTP	wastewater treatment plant

Evaluation and Use of U.S. Environmental Protection Agency Clean Watersheds Needs Survey Data to Quantify Nutrient Loads to Surface Water, 1978–2012

By Tamara Ivahnenko

Abstract

Changes in municipal and industrial point-source discharges over time have been an important factor affecting nutrient trends in many of the Nation's streams and rivers. This report documents how three U.S. Environmental Protection Agency (EPA) national datasets—the Permit Compliance System, the Integrated Compliance Information System, and the Clean Watersheds Needs Survey—were evaluated for use in the U.S. Geological Survey National Water-Quality Assessment project to assess the causes of nutrient trends. This report also describes how a database of total nitrogen load and total phosphorous load was generated for select wastewater treatment facilities in the United States based on information reported in the EPA Clean Watersheds Needs Survey. Nutrient loads were calculated for the years 1978, 1980, 1982, 1984, 1986, 1988, 1990, 1992, 1996, 2000, 2004, 2008, and 2012 based on average nitrogen and phosphorous concentrations for reported treatment levels and on annual reported flow values.

The EPA Permit Compliance System (PCS) and Integrated Compliance Information System (ICIS), which monitor point-source facility discharges, together are the Nation's most spatially comprehensive dataset for nutrients released to surface waters. However, datasets for many individual facilities are incomplete, the PCS/ICIS historical data date back only to 1989, and historical data are available for only a limited number of facilities. Additionally, inconsistencies in facility reporting make it difficult to track or identify changes in nutrient discharges over time. Previous efforts made by the U.S. Geological Survey to "fill in" gaps in the PCS/ICIS data were based on statistical methods—missing data were filled in through the use of a statistical model based on the Standard Industrial Classification code, size, and flow class of the facility and on seasonal nutrient discharges of similar facilities. This approach was used to estimate point-source loads for a single point in time; it was not evaluated for use in generating a consistent data series over time.

Another national EPA dataset that is available is the Clean Watersheds Needs Survey (CWNS), conducted every 4 years beginning 1973. The CWNS is an assessment of the capital needs of wastewater facilities to meet the water-quality goals set in the Clean Water Act. Data collected about these facilities include location and contact information for the facilities; population served; flow and treatment level of the facility; estimated capital needs to upgrade, repair, or improve facilities for water quality; and nonpoint-source best management practices.

Total nitrogen and total phosphorous load calculations for each of the CWNS years were based on treatment level information and average annual outflow (in million gallons per day) from each of the facilities that had reported it. Treatment levels categories (such as Primary, Secondary, or Advanced) were substituted with average total nitrogen and total phosphorous concentrations for each treatment level based on those reported in literature. The CWNS dataset, like the PCS/ICIS dataset, has years where facilities did not report either a treatment level or an annual average outflow, or both. To fill in the data gaps, simple linear assumptions were made based on each facility's responses to the survey in years bracketing the data gap or immediately before or after the data gap if open ended. Treatment level and flow data unique to each facility were used to complete the CWNS dataset for that facility.

Introduction

Changes in municipal and industrial point-source discharges over time have been an important factor affecting nutrient trends in many of the Nation's streams and rivers. These discharges have been monitored through the U.S. Environmental Protection Agency (EPA) Permit Compliance System (PCS) and Integrated Compliance Information System (ICIS), which combined are the Nation's most spatially comprehensive dataset for nutrients released to surface waters. The PCS, which began in 1990, was designed for tracking permits, compliance, and enforcement status for the National Pollutant Discharge Elimination System (NPDES) as part of the Clean Water Act (33 U.S.C. 1251 et seq.). The PCS dataset contains information on pollutant discharge limits and concentrations of pollutants measured in the facilities' wastewater discharges, and it tracks the facility history of compliance and construction (EPA, 1990). The EPA, to modernize the PCS dataset, implemented ICIS in 2006, which not only incorporated the data from the PCS but also included compliance data for air quality and hazardous waste facilities (EPA, 2016a). The PCS/ICIS dataset is not ideal, however, for tracking changes in nutrient discharges over time. Datasets for many individual facilities are incomplete, with as much as 10 months of data missing in a year. Moreover, the EPA PCS/ICIS historical data date back only as far as 1989 and are available for only a limited number of facilities.

Previous efforts have been made by the U.S. Geological Survey (USGS) National Water Quality Assessment (NAWQA) project to “fill in” these gaps. Missing data for a facility were filled in through the use of a statistical model based on the Standard Industrial Classification (SIC) code, size, and flow class of the facility and on the seasonal flow and nutrient discharges of the facility and (or) similar facilities (McMahon and others, 2007). This approach was used to estimate point-source loads for a single point in time; it was not evaluated for use in generating a consistent data series over time.

Another national EPA dataset that is available to assess nutrient trends is the Clean Watersheds Needs Survey (CWNS). The survey, conducted every 4 years beginning in 1973, is an assessment of the capital or financial needs of wastewater treatment facilities to meet the water-quality goals set in the Clean Water Act (EPA, 2015). Data collected about these facilities include location and contact information for the facilities; population served; flow and treatment level of the facility; estimated capital needs to upgrade, repair, or improve facilities for water quality; and nonpoint-source (NPS) best management practices. The treatment level and average annual flow data from the CWNS may be the best indicator of nutrient discharges over time. Treatment levels for wastewater are an indication of how much nitrogen and phosphorous is removed from the effluent before it is discharged back to the environment, and the survey includes data on wastewater treatment levels dating to the 1970s on a national scale. Primary treatment is designed to use physical means to remove gross, suspended, and floating solids from raw sewage, and Secondary and Advanced (tertiary) treatments use biological methods in which increasing treatment levels remove increasing amounts of nutrients from wastewaters (EPA, 2004). Changes in treatment levels in a wastewater facility can affect nutrient levels in the surface water body that receives the effluent, which can affect trends in nutrient loads. In this report, the PCS/ICIS and the CWNS datasets are evaluated for use in surface-water nutrient long-term trends models; the calculation of nutrient loads by applying literature-based (EPA, 2000) nutrient concentrations to the reported CWNS treatment levels is also described.

Purpose and Scope

The purpose of this report is to document how EPA datasets were evaluated for use in the USGS NAWQA project as a consistent time series of nutrient loading from point sources, which are a critical component of assessing the causes of nutrient trends in rivers and streams. This report also describes how a database of total nitrogen and total phosphorous loads was generated for select wastewater treatment facilities in the conterminous United States based on information reported in the EPA CWNS. Nutrient loads were calculated for the years 1978, 1980, 1982, 1984, 1986, 1988, 1990, 1992, 1996, 2000, 2004, 2008, and 2012 from the reported treatment level and annual flows for each facility. In order to use the treatment level information in the load calculations, data reported by the EPA (2000) were used to convert the reported treatment levels (primary, secondary, advanced) to numerical nitrogen and phosphorous concentrations. The final database can be found as a data release (Ivahnenko, 2017) at http://doi.org/10.5066/F7MG7MNN.

Evaluation and Use of the Data

National point-source datasets, the PCS, the ICIS, and the CWNS from the EPA, were evaluated for use in calculating an annual load of total nitrogen (TN) and total phosphorous (TP) for point-source facilities. Techniques and methods for data interpolation and “filling in” data gaps where yearly concentrations or flows were missing are discussed. Nutrient loads for each facility were calculated as the average monthly concentration of the nutrient (either nitrogen or phosphorous) multiplied by the average daily reported flow and the number of days in the month. The monthly values for each facility were then summed, and the annual nutrient load was expressed as kilograms of either TN or TP per year. The calculation for monthly nutrient loads is expressed in the following equation (TN is calculated in this example; to calculate TP, use the same equation with phosphorous concentrations in place of nitrogen concentrations):

$$TN = ((C \times 0.000001) \times (F \times 3785411.78)) \times N \quad (1)$$

where

TN	is total nitrogen, in kilograms per month;
C	is the concentration of nitrogen, in milligrams per liter;
F	is flow, in million gallons per day;
N	is the number of days in the month;
3785411.78	is a conversion factor from million gallons per day to liters per day; and
0.000001	is a conversion factor from milligrams per liter to kilograms per liter.

Evaluation of the Integrated Compliance Information System Data

The McMahon and others (2007) approach to complete the EPA PCS dataset had a ranked approach. For facilities with complete effluent-monitoring records, effluent flow and nutrient concentration data were used to calculate annual point-source nitrogen and phosphorous loads. If records contained flow but no nutrient data, a “typical pollutant concentration” (TPC) was developed based on pooled and median seasonal concentration values based on flow class, SIC code, and season of the year. These data were pooled into regional databases and used to fill in missing nutrient data for those facilities that had flow data but no corresponding nutrient data. If no TPC could be drawn from the regional TPC dataset, a national EPA dataset based on SIC codes was used to assign nutrient concentrations.

To test whether the McMahon and others (2007) approach was suitable for filling in missing point-source data in the EPA PCS and ICIS datasets for trend analysis, 4 NPDES wastewater treatment facilities (SIC 4952, Sewerage Systems) were selected at random for a detailed analysis out of the 11,399 facilities in the datasets that had complete data for the period 1990 to 2010. Total nitrogen was measured for one facility, CO0035939, and total phosphorous was measured for three facilities, AR0020010, GA0020052, and GA0021610. The observed TN and TP loads for these facilities are shown in figure 1, along with the population density of the U.S. Census blocks within a 2-mile radius of each facility (GeoLytics, 2013). The superintendents of the facilities were contacted for explanations for the changes in nutrient loads, and their comments on increases and decreases in nutrient loads are included in figure 1. For example, at facility CO0035939 (the Glacier Club at Tamarron, a golf and vacation resort in Durango, Colorado), the superintendent commented “Resort sold, number of conventions decreasing” for 1994. Other superintendents commented on when a wastewater treatment facility was upgraded or renovated, when permits changed, or when higher flow years from precipitation or the addition of industrial plants contributed effluent to the wastewater facility (such as effluent from a plating plant near facility AR0020010 in 1994 and a Kia Motor Co. plant near GA0020052 in 2009).

In order to use the complete datasets to evaluate the effectiveness of the McMahon and others (2007) approach for calculating annual loads with datasets missing nutrient values, the monthly nutrient data (nitrogen or phosphorous) for years 1992, 1997, and 2002 were removed from the complete datasets of the four selected facilities. The data were removed at monthly increments so that 25, 50, or 75 percent of the nutrient data was missing; that is, at 75 percent missing, only the data for January, June, and December remained. Facilities missing 25 percent of the monthly data or more accounted for 5–35 percent of the PCS/ICIS datasets (McMahon and others, 2007), so removing the flow and concentration data in increments represents the varying data gaps in the PCS/ICIS datasets. The McMahon and others (2007) approach was used with each of the missing data increments to estimate a TN or TP load for the facility. The results are shown in figure 2 for each year, for each facility, alongside the actual loads taken from the complete datasets.

Only in the case of facility CO0035939 did the overall temporal pattern of the estimated loads compare well with the temporal pattern of the original observed loads; that is, the three estimated loads and the observed load all followed a pattern of decrease from 1992 to 1997 and increase from 1997 to 2002. The magnitudes of the nitrogen loads, however, were poorly estimated compared to the original data. In the other cases, the overall original observed temporal pattern was not well captured by the estimated loads. Use of these estimated loads would lead to erroneous conclusions about the influence of changes in point-source loading over time. These four selected facilities are a small subset of the 11,399 facilities with complete datasets in the EPA NPDES database, but they provide an indication of the issues in using the McMahon and others (2007) approach to calculate loads for trend analysis.

An alternative, simplified method of filling in data gaps through linear interpolation was also evaluated. Nutrient and flow data for remaining months of data were substituted for the missing months. In the case where 75 percent of the monthly data was missing, the data from January were used for February, March, and April; the data from June were used for May, July, and August; and the data from December were used for September, October, and December. The nutrient loads for the same four facilities estimated by linear interpolation are shown in figure 3. As with the McMahon and others (2007) approach, in most cases, the temporal pattern of the original observed loads was not well captured by the loads estimated through interpolation.

Percent differences between the original observed nutrient loads and the loads estimated by the McMahon and others (2007) approach and by linear interpolation for the 25, 50, and 75 percent missing data increments are shown for each facility for years 1992, 1997, and 2002 in table 1. On average, the original observed loads are closer to the loads estimated by linear interpolation than to those estimated using the McMahon and others (2007) approach.

Another option evaluated for use in analyzing trends in point-source discharge was the use of population as a surrogate. As seen in figure 1 (though only for four facilities), the temporal pattern of change in population does not always correspond well to the temporal pattern of change in point-source discharge. Other influences such as increased precipitation, changes in industrial inputs, and wastewater treatment facilities renovations and upgrades have a greater yearly effect on nutrient discharge than population.

Lastly, an attempt was made to evaluate additional sources of data or additional sources of information that could be used to fill in data gaps in the PCS record. Incomplete facility datasets could be filled in by contacting the facility superintendents directly for additional data or by using the methods used by the Chesapeake Bay Program (described later). Also, some States have State records accessible online, including wastewater facility nutrient reports. For example, nutrient records for New Jersey are available through the New Jersey Department of Environmental Protection (DEP) Data Miner (New Jersey DEP, 2015), for select facilities to the year 2000.

Contacting facility superintendents individually was a time-intensive effort and yielded little or no information. The superintendents contacted indicated that the facilities often do not keep records older than 5 to 10 years, relying on the State and the EPA to keep the oldest records. According to Bruce Smith, Hackettstown Municipal Utility Authority, incomplete

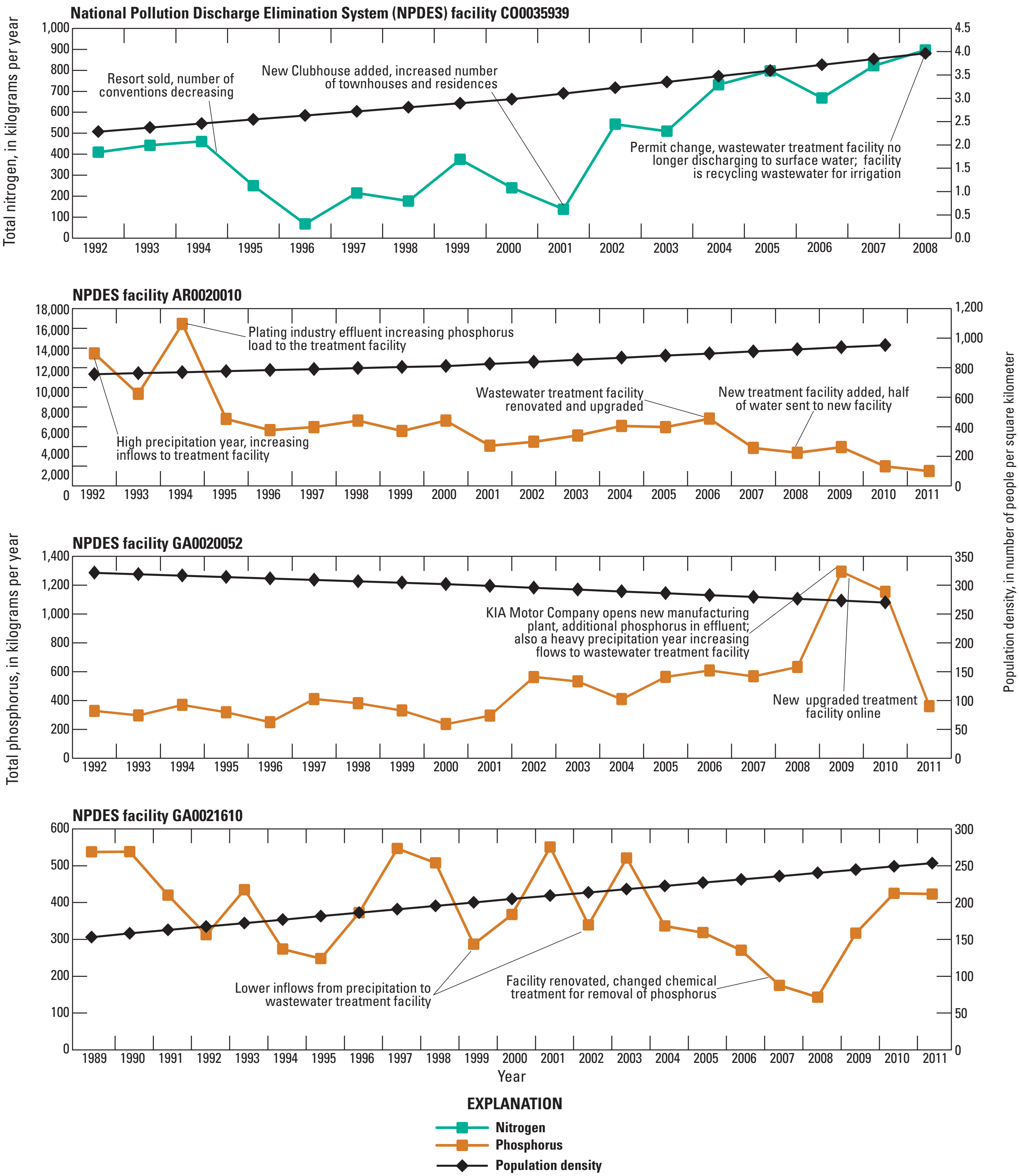

Figure 1. Annual total nitrogen and phosphorous loads for four selected National Pollution Discharge Elimination System (NPDES) facilities with complete nutrient datasets, with population density of U.S. Census blocks within a 2-mile radius of the facility. (Population data are from GeoLytics, 2013.)

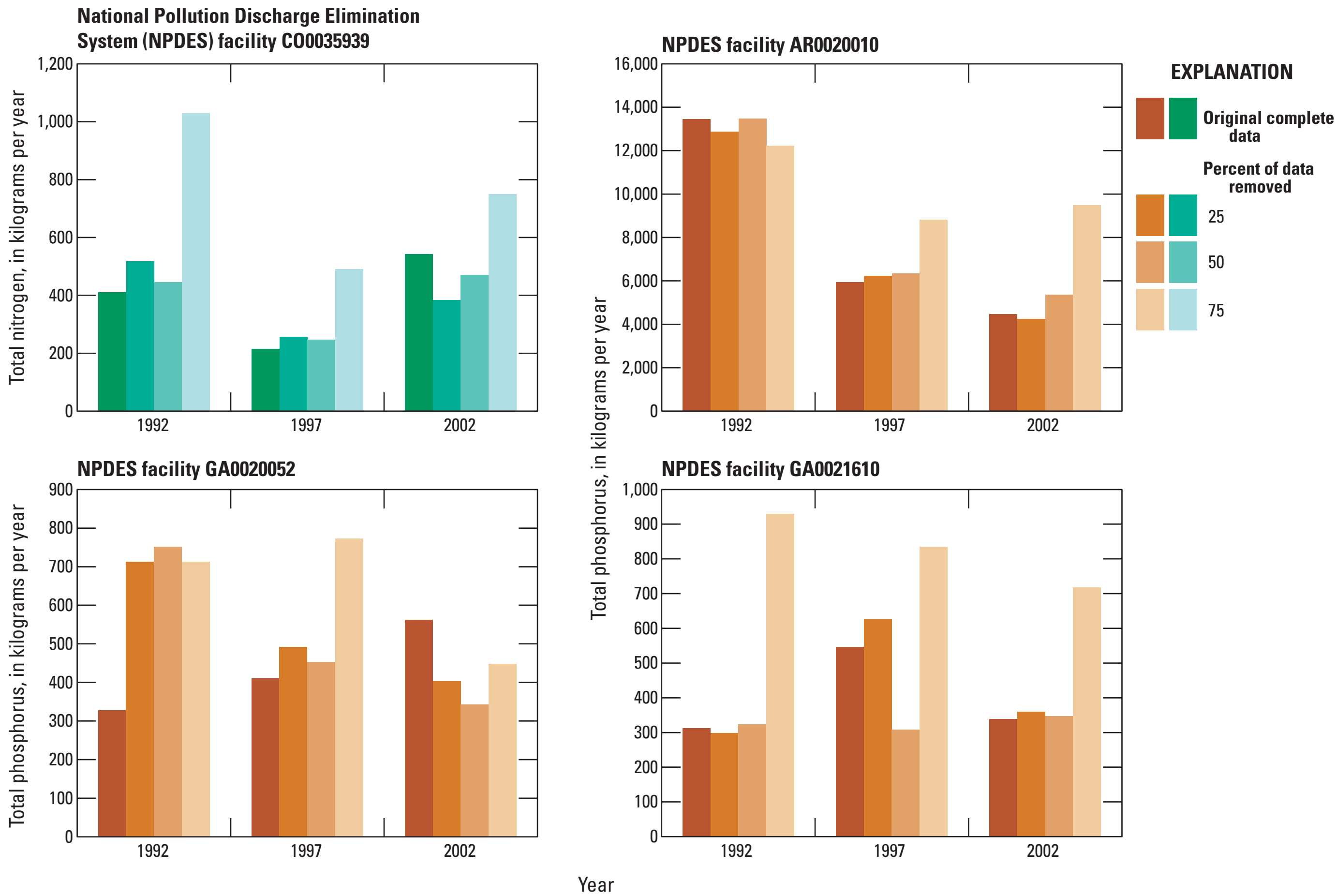

Figure 2. Annual total nitrogen and phosphorous loads calculated by the McMahon and others (2007) approach for complete datasets with 25, 50, and 75 percent of data removed. Four National Pollution Discharge Elimination System (NPDES) facilities with complete nutrient datasets for 1992, 1997, and 2002 were selected. Actual nutrient loads from the complete datasets shown for comparison.

data from the year 1992 were the oldest effluent data available through individual superintendent contacts. Even if contacting individual wastewater facilities yielded useful information, doing so for the thousands of sites in the dataset was clearly out of the scope of effort for this project.

The Chesapeake Bay Program website has links to water-quality, biological, point-source, modeling, and geographic information system (GIS) datasets (Chesapeake Bay Program, 2015). The current (2014) Chesapeake Bay Program policy on filling in missing point-source data is to rely on the data providers to fill in any data gaps; the methods used to estimate missing loads vary amongst the States that provide the data (EPA, 2016b). The program’s method for filling in missing data could not be used nationally for PCS/ICIS data gaps because it would require all the State agencies to develop default nutrient values, which is not mandatory outside of the Chesapeake Bay watershed. The method used by the Chesapeake Bay Program Office was not feasible to use for the NAWQA project effort for assessing the causes of nutrient trends.

In investigating the use of State data pages, such as the New Jersey DEP Data Miner, it was found that electronic data are available for at most a 10-year historical record. Some State agencies had microfilm and microfiche records older than the 10-year historical records online, but they would have to be accessed and procured in person. In other states such as Pennsylvania and New York, electronic records were only available for a 5-year period, and again, an individual would have to access the microfilm or paper records in person to complete the historical nutrient and effluent dataset for the facilities in those states. As with contacting each wastewater facility superintendent, this effort would be time intensive and would not yield a dataset in time for inclusion in the nutrient trends model currently (2016) being developed.

Evaluation of the Clean Watershed Needs Survey Data

The EPA CWNS is primarily designed to evaluate the need of wastewater treatment facilities to upgrade or replace existing facilities and infrastructure to meet State and local water-quality guidelines in response to the Federal Clean

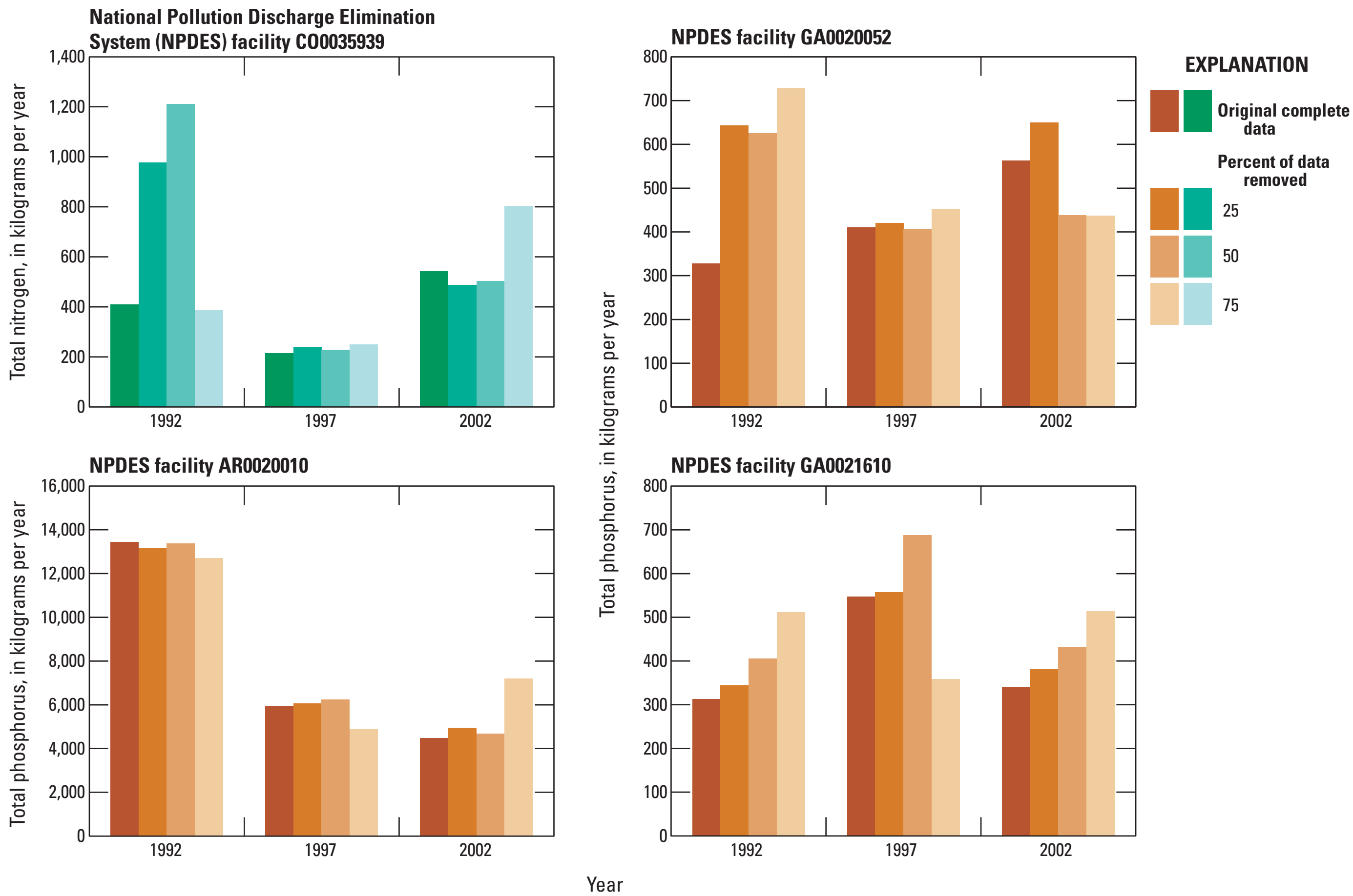

Figure 3. Annual total nitrogen and phosphorous loads calculated by linear interpolation for complete datasets with 25, 50, and 75 percent of data removed. Four National Pollutant Discharge Elimination System (NPDES) facilities with complete nutrient datasets for 1992, 1997, and 2002 were selected. Actual nutrient loads from the complete datasets shown for comparison.

Water Act, Sections 205(a) and 516 (EPA, 2015). Facilities voluntarily respond to the survey; the information provided can be used to apply for grants and is sent to Congress and State legislatures to assist in creating budgets for improving and upgrading treatment facilities. The initial CWNS was conducted in 1973 and was repeated in 1976. Data from these two surveys are in different categories and different formats from the data reported in surveys conducted in 1978 and later, so they were not incorporated into the nutrient load database (Ivahnenko, 2017). Minor rather than major (more than 1 million gallons per day discharged) facilities are better represented in the CWNS dataset (fig. 4) because, nationally, there are more minor facilities than major. Based on geographic region, Maupin and Ivahnenko (2011) reported minor facilities represented 83 to 97 percent of all the facilities in the 2002 EPA PCS/ICIS dataset.

A section of the CWNS is devoted to information on the facility, such as location, treatment level, population served, and annual average outflow. The discharge locations reported in the CWNS varied in precision and accuracy. To ensure the reported discharge locations were as accurate as possible, the reported location was mapped in ArcGIS (Esri, 2016) to confirm the reported State, county, and watershed information. When there was a clear discrepancy—the reported location was not on a stream or was in the wrong State, county, or watershed—further research was conducted. This research sometimes necessitated contact with facility operators or examination of locations in State databases. If street addresses were known, locations were obtained by using Google Earth™. Ultimately, the locations of 778 facilities were updated through this process and added to the CWNS dataset. For facilities that still had missing location information, latitude and longitude were determined by matching the facility name to a city or town in the same State or county and using the information for that city or town or by using the centroid location of the city or town (Ivahnenko, 2017). Treatment levels used at the facilities, as reported in the CWNS, are shown by year in figure 5. Raw discharge (no treatment) was not reported after 1992, and Primary and Advanced Primary treatment declined to zero after 1996 (fig. 5). Secondary water

Table 1. Percent difference between original observed total nitrogen and total phosphorous loads and loads estimated by the McMahon and others (2007)[1] approach and by linear interpolation for datasets with missing data. Original nutrient load data from four selected National Pollutant Discharge Elimination System (NPDES) facilities with complete nutrient datasets for years 1992, 1997, and 2002. Data were removed at increments of 25, 50, and 75 percent to simulate missing data.

[TN, total nitrogen; TP, total phosphorous; kg, kilogram; NPDES, National Pollutant Discharge and Elimination System. Nine-digit identification codes are given for the NPDES facilities]

Removed data, in percent	TN in kg NPDES facility CO0035939		TP in kg NPDES facility GA0021610		TP in kg NPDES facility GA0020052		TP in kg NPDES facility AR0020010	
	Percent difference, McMahon	Percent difference, linear	Percent difference, McMahon	Percent difference, linear	Percent difference, McMahon	Percent difference, linear	Percent difference, McMahon	Percent difference, linear
				1992				
25	23.4	81.8	4.8	9.5	74.0	65.0	4.4	2.0
50	8.6	99.0	3.3	25.8	78.5	62.5	0.2	0.6
75	86.1	5.7	99.4	48.4	74.0	75.8	9.5	5.7
				1997				
25	17.6	11.0	13.5	1.9	18.1	2.5	4.7	1.7
50	13.6	6.1	55.8	22.9	9.9	1.1	6.4	4.6
75	78.1	15.4	41.7	41.7	61.2	9.6	38.7	20.0
				2002				
25	34.2	10.6	5.6	11.4	33.0	14.3	5.0	10.0
50	14.3	7.4	5.0	23.8	48.7	25.0	18.0	4.7
75	32.1	38.8	71.6	40.8	22.6	25.3	71.8	46.8

[1]McMahon, Gerard; Tervelt, Larinda; and Donehoo, William, 2007, Methods for estimating annual wastewater nutrient loads in the southeastern United States: U.S. Geological Survey Open-File Report 2007–1040, 81 p.

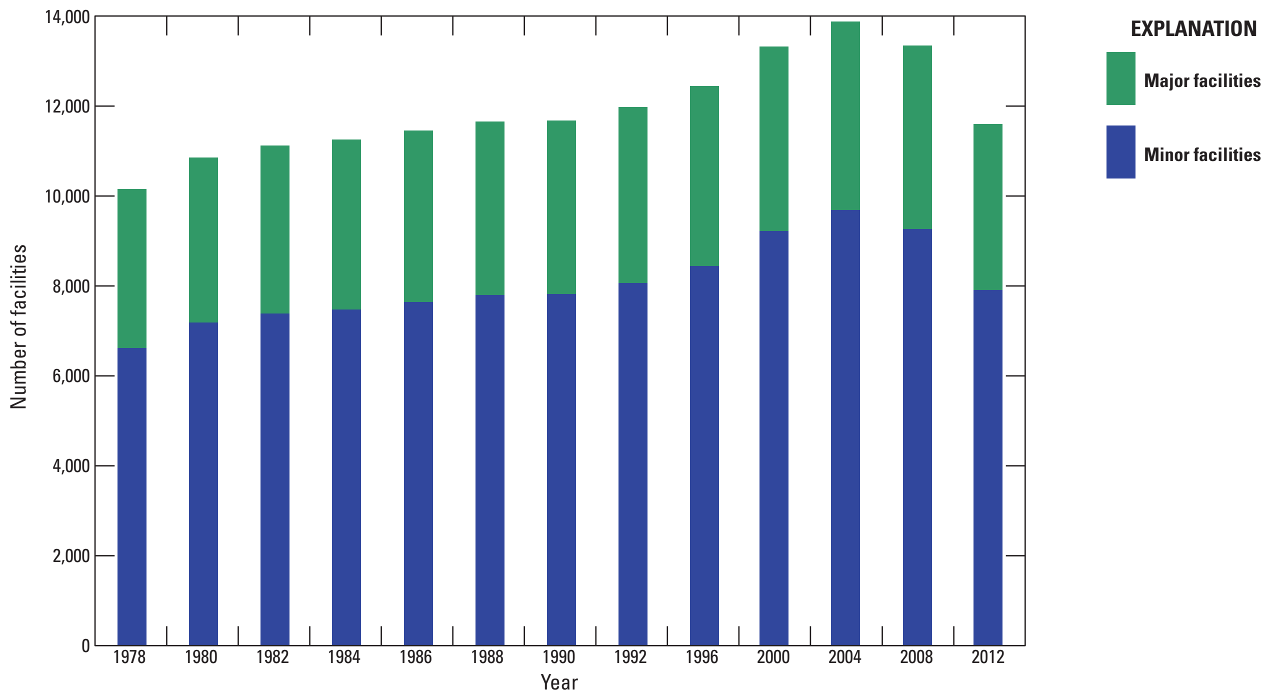

Figure 4. Distribution of major and minor wastewater treatment facilities from the Clean Watersheds Needs Survey, 1978 to 2012. Major facilities discharge more than 1 million gallons per day.

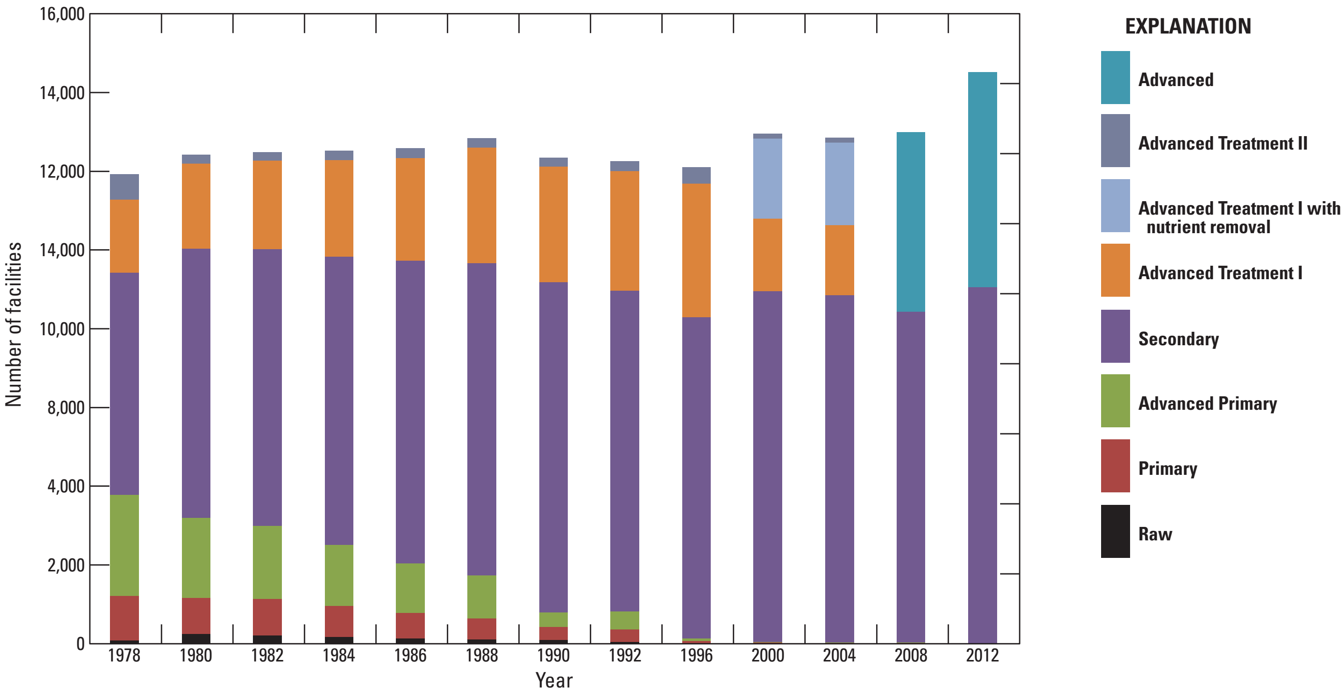

Figure 5. Distribution of reported wastewater treatment levels from facilities in the Clean Watersheds Needs Survey, 1978 to 2012.

treatment ranged from 46 percent of the reported facilities in 1978 to 67 percent of the facilities in 2000, and since 1990, Secondary treatment has constituted 60 percent or more of the treatment levels reported in the CWNS. Prior to the 2008 survey, the EPA used the treatment categories Advanced Treatment I, Advanced Treatment I with nutrient removal, and Advanced Treatment II, but the EPA combined those categories for the 2008 and 2012 surveys to Advanced.

Ideally, for the nutrient models being developed, the CWNS dataset would have the same number of facilities as the national PCS/ICIS dataset. This would provide national coverage of TN and TP loads over the span of the CWNS in the same surface water body. The number of facilities in the 2012 CWNS dataset was compared to the number of facilities in the 2012 ICIS dataset by state and NPDES identification number. The CWNS dataset is complete with regard to the ICIS dataset (100 percent of ICIS facilities also in the CWNS dataset) in Washington D.C. and 14 states (fig. 6). The CWNS is 75 percent complete with regard to the ICIS dataset in an additional 16 states (fig. 6). Only Minnesota, South Dakota, and Wyoming have less than 25 percent of the facilities in common between the CWNS and the ICIS. South Carolina did not have any facilities participate in the 2012 CWNS.

The CWNS is a more ideal dataset than the PCS/ICIS for use as a potential explanatory variable in assessing nutrient trends. Although the PCS/ICIS dataset has measured nutrient concentrations and effluent volumes, the dataset only goes back to 1990 for a small number of facilities and, for many facilities, the data record is incomplete, requiring estimated data or archived original data to fill in the gaps. No suitable estimation method can recreate the missing records, and archived data are difficult to obtain. The CWNS dataset, based on treatment level information and annual flow volumes and with information for a number of facilities dating back to 1978, will provide more information on the changes in nutrient values in surface water for a longer period of record.

Calculation of Total Nitrogen and Phosphorous Loads from the Clean Watershed Needs Survey

Calculations for TN and TP loads for each year of the CWNS were based on treatment level information and average annual outflow data in million gallons per day from each facility that reported it. The treatment levels reported in the CWNS were converted to numerical average concentrations of TN and TP according to data reported by the EPA (2000) (for example, the Primary treatment level equates to an average TN concentration of 23.4 milligrams per liter); the substitutions are listed in table 2. The average nutrient concentrations reported by the EPA (2000) and listed in table 2 were generated from several sources, primarily nitrogen and phosphorous concentrations reported in the EPA PCS dataset, the CWNS dataset, and other literature. For surveys that used the treatment level categories Advanced Treatment I, Advanced Treatment I with nutrient removal, and Advanced Treatment II (CWNS years prior to 2000), the average concentrations of TN and TP for the single Advanced category were applied to the three replaced "subcategories" in order to calculate nutrient loads in the nutrient model because the

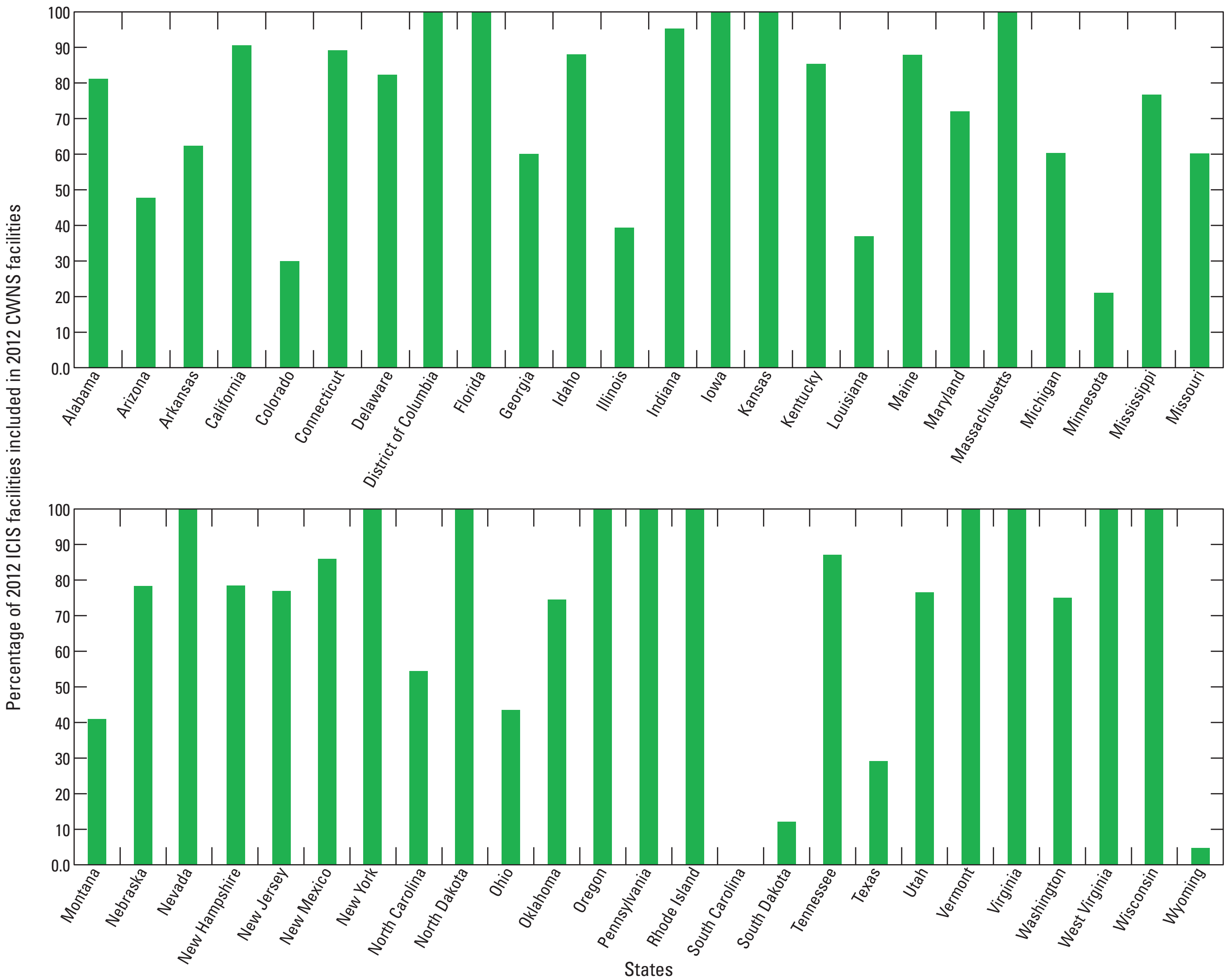

Figure 6. Comparison of the number of facilities in both the 2012 Clean Watershed Needs Survey (CWNS) dataset and the 2012 Integrated Compliance Information System (ICIS) dataset.

Table 2. Total nitrogen and total phosphorous concentrations of effluent used in Clean Watershed Needs Survey load calculations.

[mg/L, milligram per liter]

Treatment	Nutrient concentrations[1]	
	Total nitrogen, in mg/L	Total phosphorous, in mg/L
Raw discharge	30	6
Primary	23.4	5.2
Advanced primary	23.4	5.2
Secondary	18.3	2.5
Advanced secondary	18.4	0.4
Advanced treatment	14.4	0.4

[1]Concentrations reported in U.S. Environmental Protection Agency (2000).

subcategories were not used in EPA (2000), only the single Advanced category. The average nutrient concentrations reported in EPA (2000) for the Advanced category might overestimate the actual nutrient concentrations associated with the three subcategories, but there is no evidence in the CWNS dataset that facilities downgraded from an advanced treatment to a secondary or lower level of treatment (fig. 5), so any overestimates for nutrient concentrations and loadings from the simplified Advanced category would be consistent and insignificant through all the years of the dataset, as the three subcategories were used in all years of the CWNS from 1978 to 2004, when they were simplified by the EPA.

The CWNS dataset, like the PCS/ICIS dataset, has gaps in the record for some facilities when the facility did not report either a treatment level or an annual average outflow. In the CWNS dataset, the gaps were annual values missing from the record; in the PCS/ICIS dataset, the gaps were seasonal or monthly values. Linear interpolation was used to fill in the annual gaps in the CWNS dataset. The finer scaled seasonal and monthly dataset in McMahon and others (2007), however, incorporates more variability in seasonal and monthly treatment and discharge practices for similar facilities, SIC codes, and flow magnitudes in the calculation of annual loads. As a result, linear interpolation of missing PCS/ICIS seasonal and monthly values may not be as appropriate as linear interpolation of missing CWNS annual values.

Linear interpolation of nutrient loads for a facility in the CWNS dataset requires annual treatment level and flow data from that facility reported for the years bracketing the data gap or for a year on one end of the data gap. For example, if a facility had a missing treatment level or did not report a treatment level for a number of years but had reported using a Secondary treatment level before and after the data gap, then it was assumed that treatment levels had not changed for the missing years of data. A treatment level of Secondary was applied to the years with missing data and a load was calculated. The same assumption would apply to missing annual outflow volumes. In the case that different treatment levels or outflow volumes were reported in surveys before and after the data gap, then the number of missing data points would be divided equally between the different treatment levels and outflow volumes. For example, if a facility reported a treatment level of Secondary in an early survey then reported a treatment level of Advanced after missing three intervening surveys, then the missing data for the three survey years would be filled in with data for two Secondary treatments and outflow volumes and one Advanced treatment and outflow volume. Finally, for open-ended data gaps, meaning that the years of missing data are not bracketed by reported treatment level or outflow data, then the values for the first reported year adjacent to the data gap of treatment level and outflow were extended to all the missing years of data. About 89 percent of the facility records had at least one year of missing CWNS data filled in using linear interpolation. In 2012, none of the facilities in South Carolina submitted a survey to the CWNS, so they are unrepresented in the database for that year. Unlike the McMahon and others (2007) method used to complete the PCS/ICIS dataset, in which data from similar facilities were used to fill the data gaps, this method used treatment level and flow data unique to each facility to complete the CWNS dataset. The final dataset can be found as a data release (Ivahnenko, 2017; http://doi.org/10.5066/F7MG7MNN).

Summary

Changes in municipal and industrial point-source discharges over time have been an important factor affecting nutrient trends in many of the Nation's streams and rivers. These discharges have been monitored through the U.S. Environmental Protection Agency (EPA) Permit Compliance System (PCS) and Integrated Compliance Information System (ICIS), which combined are the Nation's most spatially comprehensive dataset for nutrients released to surface waters. The PCS/ICIS historical data date back only to 1989 and are available for only a limited number of facilities. Additionally, nutrient and flow datasets for many individual facilities are incomplete. To track changes in nutrient discharges over time, this dataset in its unaltered form is not ideal. Previous efforts were made by the USGS to "fill in" gaps in the PCS/ICIS data. In the previous USGS nutrient load calculation approach, missing data for a facility were filled in through the use of a statistical model based on the facility's Standard Industrial Classification (SIC) code, size, and flow class and on the seasonal nutrient discharges of similar facilities. This approach was used to estimate point-source loads for a single point in time; it was not evaluated for use in generating a consistent data series over time.

This report documents how three EPA national datasets—PCS, the ICIS, and the Clean Watersheds Needs Survey (CWNS)—were evaluated for use in the U.S. Geological Survey (USGS) National Water Quality Assessment (NAWQA) project to assess the causes of nutrient trends. This report also describes how a database of total nitrogen load and total phosphorous load was generated for select wastewater treatment facilities in the conterminous United States based on information reported in the EPA CWNS. Nutrient loads for total nitrogen (TN) and total phosphorous (TP) were calculated for the years 1978, 1980, 1982, 1984, 1986, 1988, 1990, 1992, 1996, 2000, 2004, 2008, and 2012 based on average nitrogen and phosphorous concentrations for reported treatment levels and on annual reported flow values.

To test whether the previous USGS approach was suitable for filling in missing point-source data in the EPA PCS and ICIS datasets for trend analysis, four National Pollutant Discharge Elimination System wastewater treatment facilities with complete datasets for the period 1990 to 2010 were selected from the 11,399 facilities included in the EPA datasets. Data were removed from the datasets in increments so that 25, 50, or 75 percent of the nutrient data was missing to introduce data gaps similar to those in the EPA datasets, until only the data for January, June, and December remained. The statistical model from the previous USGS approach, based on a facility's SIC code and flow class and on the seasonal nutrient discharges of similar facilities, was used with each of the missing data increments to estimate a TN or TP load for the facility. Only in a single case did the temporal pattern of loads estimated for missing data compare well with the temporal pattern of the original observed loads over all analyzed years. In the other cases, the original observed temporal pattern was not well captured by the estimated loads.

As an alternative to the statistical model used in the previous USGS approach, a simplified method of filling in data gaps using linear interpolation was evaluated, again using data from four facilities with complete flow and nutrient records for all years during 1990 to 2011. As with the previous USGS approach, the original observed temporal pattern was not well captured by the loads estimated through interpolation. Using the same four facilities, another option, the use of population as a surrogate, was evaluated for use in analyzing trends in point-source discharge. There was a poor correspondence between the temporal patterns of population change and point-source discharge. Other influences such as increased precipitation, changes in industrial inputs, and wastewater treatment facility renovations and upgrades have a greater yearly effect on nutrient discharge than population.

The ability to obtain the missing data from other sources was also explored. Contacting facility superintendents individually indicated that the facilities often do not keep records older than 5 to 10 years, relying on the State and the EPA to keep the oldest records. State agencies responsible for monitoring wastewater treatment facility effluent also do not archive water quality and outflow volumes for more than 10 years.

Another national EPA dataset that is available for use is the CWNS. The survey, conducted every 4 years beginning in 1973, is an assessment of the capital needs of wastewater facilities to meet the water-quality goals set in the Clean Water Act. Data collected about these facilities include location and contact information for the facilities; population served; flow and treatment level of the facility; estimated capital needs to upgrade, repair, or improve facilities for water quality; and nonpoint-source best management practices. The CWNS dataset, based on treatment level information and annual flow volumes, includes information for a number of facilities dating back to 1978, so it will provide more information on the changes in nutrient values in surface water for a longer period of record.

Treatment level information and average annual outflow (in million gallons per day) from each facility that reported it were the basis for the TN and TP load calculations for each of the CWNS years. The treatment levels were converted to numerical average TN and TP concentrations based on those reported in literature. The CWNS dataset, like the PCS/ICIS dataset, has years when the facility did not report a treatment level, an annual average outflow, or both. For open-ended data gaps, meaning that the years of missing data are not bracketed by reported treatment level or outflow data, then the values for the first reported year adjacent to the data gap of treatment level and outflow were extended to all the missing years of data. To fill in the data gaps, linear interpolations were made based on each facility's responses to the survey in the years bracketing the data gap or immediately before or after the data gap if open-ended. Unlike the previous USGS approach used to complete the PCS/ICIS dataset, in which data from similar facilities were used to fill the data gaps, this method used treatment level and flow data unique to each facility to complete the CWNS dataset.

References Cited

Chesapeake Bay Program, 2015, Data hub: Chesapeake Bay Program web page, accessed August 12, 2015, at http://www.chesapeakebay.net/data.

Esri, 2016, ArcGIS for Desktop, release 10.3.1: Redlands, Calif., Environmental Systems Research Institute.

GeoLytics, 2013, Census time series research package—2010 summary file (SF1), 2000 in 2010 boundaries, and 1990 in 2010 boundaries: East Brunswick, N.J., GeoLytics, Inc., 3 CD-ROMs.

Ivahnenko, Tamara, 2017, National EPA Clean Watershed Needs Survey WWTP nutrient loads 1978 to 2012: U.S. Geological Survey data release, accessed September 25, 2017, at https://doi.org/10.5066/F7MG7MNN.

Maupin, M.A., and Ivahnenko, Tamara, 2011, Nutrient loadings to streams of the continental United States from municipal and industrial effluent: Journal of the American Water Resources Association, v. 47, no. 5, p. 950–964.

McMahon, Gerard; Tervelt, Larinda; and Donehoo, William, 2007, Methods for estimating annual wastewater nutrient loads in the southeastern United States: U.S. Geological Survey Open-File Report 2007–1040, 81 p.

New Jersey Department of Environmental Protection [DEP], 2015, DEP data miner: New Jersey Department of Environmental Protection web page, accessed October 2, 2017, at https://www13.state.nj.us/DataMiner/Search/SearchByCategory?isExternal=y&getCategory=y&catName=Site+Remediation.

U.S. Environmental Protection Agency [EPA], 1990, Permit Compliance System: U.S. Environmental Protection Agency, 20W-4001, 21 p.

U.S. Environmental Protection Agency [EPA], 2000, Progress in water quality—An evaluation of the national investment in municipal wastewater treatment: U.S. Environmental Protection Agency, EPA-832-R-00-008, 451 p.

U.S. Environmental Protection Agency [EPA], 2004, Primer for municipal wastewater treatment systems: U.S. Environmental Protection Agency, EPA 832-R-04-001, 30 p.

U.S. Environmental Protection Agency [EPA], 2015, Clean Watersheds Needs Survey overview: U.S. Environmental Protection Agency web page, accessed August 12, 2015, at http://water.epa.gov/scitech/datait/databases/cwns/.

U.S. Environmental Protection Agency [EPA], 2016a, ECHO, Enforcement and Compliance History Online. About the data—Data completeness: U.S. Environmental Protection Agency web page, accessed December 6, 2016, at https://echo.epa.gov/resources/echo-data/about-the-data#completeness.

U.S. Environmental Protection Agency [EPA], 2016b, Chesapeake Bay Program Wastewater Facility and BMP Implementation Data Submission Specifications and Requirements; Attachment 6, accessed March 1, 2017, at https://www.epa.gov/sites/production/files/2016-01/documents/attachment6pointnonpointsourcedata.pdf

Publishing support provided by
Denver Publishing Service Center

For more information concerning this publication, contact
Program Coordinator, National Water Quality Program
U.S. Geological Survey
413 National Center
12201 Sunrise Valley Drive
Reston, Virginia 20192

Or visit the National Water Quality Program website at
https://water.usgs.gov/nawqa

USGS

Ivahnenko—**Evaluation and Use of EPA Clean Watersheds Needs Survey Data to Quantify Nutrient Loads to Surface Water, 1978–2012**—SIR 2017–5115

ISSN 2328-0328 (online)
https://doi.org/10.3133/sir20175115