

Prepared in cooperation with the U.S. Environmental Protection Agency
Chesapeake Bay Program

Application of a Weighted Regression Model for Reporting Nutrient and Sediment Concentrations, Fluxes, and Trends in Concentration and Flux for the Chesapeake Bay Nontidal Water-Quality Monitoring Network, Results Through Water Year 2012



Scientific Investigations Report 2015–5133

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

Cover. Great Falls of the Potomac River, on the Maryland-Virginia border near Washington, D.C., on November 24, 2012. The Potomac River is the second largest nontidal tributary to the Chesapeake Bay. Photograph from the National Park Service at <http://www.nps.gov/media/photo/gallery.htm?id=1D22959F-1DD8-B71C-0799879456F88177>.

Application of a Weighted Regression Model for Reporting Nutrient and Sediment Concentrations, Fluxes, and Trends in Concentration and Flux for the Chesapeake Bay Nontidal Water-Quality Monitoring Network, Results Through Water Year 2012

By Jeffrey G. Chanat, Douglas L. Moyer, Joel D. Blomquist, Kenneth E. Hyer, and Michael J. Langland

Prepared in cooperation with the U.S. Environmental Protection Agency
Chesapeake Bay Program

Scientific Investigations Report 2015–5133

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

U.S. Department of the Interior

SALLY JEWELL, Secretary

U.S. Geological Survey

Suzette M. Kimball, Acting Director

U.S. Geological Survey, Reston, Virginia: 2016

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 <http://www.usgs.gov> or call 1–888–ASK–USGS.

For an overview of USGS information products, including maps, imagery, and publications, visit <http://www.usgs.gov/pubprod/>.

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Although this information product, for the most part, is in the public domain, it also may contain copyrighted materials as noted in the text. Permission to reproduce copyrighted items must be secured from the copyright owner.

Suggested citation:

Chanat, J.G., Moyer, D.L., Blomquist, J.D., Hyer, K.E., and Langland, M.J., 2016, Application of a weighted regression model for reporting nutrient and sediment concentrations, fluxes, and trends in concentration and flux for the Chesapeake Bay Nontidal Water-Quality Monitoring Network, results through water year 2012: U.S. Geological Survey Scientific Investigations Report 2015–5133, 76 p., <http://dx.doi.org/10.3133/sir20155133>.

ISSN 2328-031X (print)

ISSN 2328-0328 (online)

ISBN 978-1-4113-4005-3

Contents

Acknowledgments.....	viii
Abstract.....	1
Introduction.....	2
Purpose and Scope	9
Technical Considerations Leading to the Adoption of WRTDS.....	10
Overview of ESTIMATOR and WRTDS Models.....	10
Previously Published Comparisons of ESTIMATOR and WRTDS	13
Comparison of ESTIMATOR and WRTDS Models on Chesapeake Bay Nontidal Water-Quality Monitoring Network Data Through Water Year 2012	14
Description of Scenarios Considered	14
Measures of Model Performance.....	15
Explanatory Power	15
Flux Bias	16
Estimated Yields.....	19
Trends in Concentration.....	20
Reconciling the Concepts of “Flow-Adjusted Concentration” and “Flow-Normalized Concentration”	24
Trends in Flux.....	29
Sensitivity of Trends to Incremental Incorporation of New Data	32
Summary of Model Comparison Findings.....	38
WRTDS Estimated Flux and Trend Results for Nutrients and Sediment for the Chesapeake Bay Nontidal Water-Quality Monitoring Network Through Water Year 2012	38
Fluxes and Yields.....	38
Trends	41
Period-of-Record Trends	41
Ten-Year Trends.....	43
Selected Technical Considerations Associated With Operational Deployment of WRTDS Across the Chesapeake Bay Nontidal Water-Quality Monitoring Network.....	45
Dense Sub-Daily Sampling.....	45
Variability in Annual Sampling Effort and Storm Sampling Effort	47
Description of Methods	51
Results: Variability in Annual Sampling Effort.....	53
Results: Variability in Storm Sampling Effort.....	59
Changes in Laboratory Methods.....	67
Summary.....	71
References Cited.....	73

Appendixes (available at <http://dx.doi.org/10.3133/sir2015513>)

1. Effects of the incremental annual incorporation of new water-quality observations on trends—Graphic results for the six watersheds subject to detailed study
2. Tabular WRTDS results for period-of-record and 10-year flow-normalized trends in concentration and flux
3. Tabular WRTDS results for all stations and constituents modeled through water year 2012 [Also available on CD-ROM]
4. Effects of variability in annual sampling effort on trends—Graphic results for the eight watersheds subject to detailed study
5. Effects of variability in storm sampling effort on trends—Graphic results for the eight watersheds subject to detailed study

Figures

1. Map showing locations of major drainage basins in the Chesapeake Bay watershed and locations of CBNTN stations for which flux and (or) trend estimates were published through water year 2012.....3
2. Histogram of water-quality record lengths for all CBNTN stations active as of fall 20149
3. Boxplots comparing model explanatory power between ESTIMATOR and WRTDS models for the three CBNTN scenarios for which WRTDS flux results were reported in water year 2012.....17
4. Boxplots comparing flux bias statistics between ESTIMATOR and WRTDS models for the three CBNTN scenarios for which WRTDS flux results were reported in water year 2012.....18
5. Scatter plots comparing estimated 5-year (water year 2008–12) average yields of nutrients and sediment computed using ESTIMATOR and WRTDS for the CBNTN 2012 *LONG_TERM* scenario.....19
6. Plots comparing the flow-adjusted trend in concentration computed using ESTIMATOR with the flow-normalized trend in concentration computed using WRTDS for the CBNTN 2012 *LONG_TERM* scenario.....21
7. Plots comparing the flow-adjusted trend in concentration computed using ESTIMATOR with the flow-normalized trend in concentration computed using WRTDS for the CBNTN 2012 *10_YEAR* scenario23
- 8–15. Graphs showing—
 8. Total phosphorus concentration observations in relation to date at the Potomac River at Chain Bridge, at Washington, DC, for water years 1985–2012 ...25
 9. Flow-adjusted trend in concentration of total phosphorus at Potomac River at Chain Bridge, at Washington, DC, for water years 1985–2012 as computed by the ESTIMATOR model applied to the data shown in figure 825
 10. Estimated concentration surfaces as functions of time and discharge for total phosphorus at the Potomac River at Chain Bridge, at Washington, DC, based on the data shown in figure 826
 11. Period-of-record trends in concentration of total phosphorus at Potomac River at Chain Bridge, at Washington, DC, comparing flow-normalized trend computed using WRTDS with flow-adjusted trend computed using ESTIMATOR28

12.	Period-of-record trends in concentration of total phosphorus at the nine Chesapeake Bay River Input Monitoring stations, comparing flow-normalized trends computed using WRTDS with flow-adjusted trends computed using ESTIMATOR	31
13.	Trends in flow-normalized concentration of total phosphorus at Potomac River at Chain Bridge, at Washington, DC, estimated using WRTDS with records beginning in 1985 and ending in sequential years from 1994 through 2012.	32
14.	Trends in flow-adjusted concentration of total phosphorus at Potomac River at Chain Bridge, at Washington, DC, estimated using ESTIMATOR with records beginning in 1985 and ending in sequential years from 1994 through 2012.....	33
15.	Graph showing the evolution in estimated values of flow-normalized concentration of total phosphorus at Potomac River at Chain Bridge, at Washington, DC, for 1994, 1997, 2000, and 2003.....	35
16.	Boxplots showing the deviation of both WRTDS and ESTIMATOR estimates of flow-normalized concentration of total phosphorus for 1994, 1997, 2000, and 2003 at three CBNTN stations from the values estimated with retrospective data for 9 additional years.....	36
17.	Boxplots showing the deviation of both WRTDS and ESTIMATOR estimates of flow-normalized concentration of dissolved inorganic nitrogen for 1994, 1997, 2000, and 2003 at three CBNTN stations from the values estimated with retrospective data for 9 additional years.....	37
18.	Plots comparing point estimates of period-of-record trends in flow-normalized concentration with corresponding trends in flow-normalized flux, in percent, for stations modeled under the <i>LONG_TERM</i> scenario through 2012.....	42
19.	Plots comparing point estimates of 10-year trends in flow-normalized concentration with corresponding trends in flow-normalized flux, for stations modeled under the <i>10_YEAR</i> scenario through 2012.....	44
20–36.	Graphs showing—	
20.	CBNTN long-term stations, ranked in decreasing order of the percentage of total water-quality samples collected in excess of one sample per day	46
21.	Plot of dissolved inorganic nitrogen in relation to date for the Conestoga River at Conestoga, PA, highlighting instances of sub-daily sampling.....	47
22.	CBNTN long-term monitoring stations ranked in decreasing order of average annual sampling frequency.....	48
23.	Number of samples collected per year at 8 of the 10 CBNTN long-term stations with the highest average annual sampling frequency.....	49
24.	Number of samples collected per year at the eight CBNTN long-term stations with the lowest average annual sampling frequency	50
25.	Comparison of overall annual sampling frequency and storm sampling frequency of a single subsample of the record for total phosphorus at the Susquehanna River at Marietta, PA, with corresponding values in the baseline dataset.....	54
26.	Storm sampling frequency summary statistics for all eight long-term stations used in the subsampling experiment.....	55
27.	Long-term trends in flow-normalized concentration and flux of dissolved inorganic nitrogen at Susquehanna River at Marietta, PA, estimated using both the baseline record and 10 “design guideline” subsamples	56
28.	Long-term trends in flow-normalized concentration and flux of total phosphorus at Susquehanna River at Marietta, PA, estimated using both the baseline record and 10 “design guideline” subsamples	57

29. Long-term trends in flow-normalized concentration and flux of sediment at Susquehanna River at Marietta, PA, estimated using both the baseline record and 10 “design guideline” subsamples	58
30. Long-term trends in flow-normalized concentration and flux of dissolved inorganic nitrogen at Susquehanna River at Marietta, PA, estimated using 10 “design guideline” subsamples, 10 “heterogeneous” subsamples, and 10 subsamples with the simulation period truncated to 2003–12	60
31. Long-term trends in flow-normalized concentration and flux of total phosphorus at Susquehanna River at Marietta, PA, estimated using 10 “design guideline” subsamples, 10 “heterogeneous” subsamples, and 10 subsamples with the simulation period truncated to 2003–12	61
32. Long-term trends in flow-normalized concentration and flux of sediment at Susquehanna River at Marietta, PA, estimated using 10 “design guideline” subsamples, 10 “heterogeneous” subsamples, and 10 subsamples with the simulation period truncated to 2003–12	62
33. Plots of observed concentration of orthophosphate over time at two long-term stations, illustrating inconsistent censoring	68
34. Plots of observed concentration of total phosphorus over time at the Conestoga River at Conestoga, PA, illustrating artificially imposed censoring conditions.....	69
35. Trends in flow-normalized concentration of total phosphorus at the Conestoga River at Conestoga, PA, contrasting trends estimated using the three datasets illustrated in figure 34.....	70
36. Plot of observed concentration of total phosphorus over time at the Choptank River near Greensboro, MD, with superimposed trends in flow-normalized concentration, illustrating the effects of a change in laboratory methods in January 1999.....	70

Tables

1. Names, physical characteristics, period of record, and model selection for CBNTN stations for which flux and (or) trend estimates were published through water year 2012	5
2. Names, physical characteristics, and period of record for CBNTN stations active as of fall 2014	8
3. Water year 2012 fluxes and 5-year (2008–12) mean yields for total nitrogen, total phosphorus, and sediment at the CBNTN stations with record lengths of at least 5 years	39
4. Subsampling algorithm parameters for each of the eight long-term stations used in the subsampling experiment	52
5. Ten-year (2003–12) trend slopes for flow-normalized concentration and flux of dissolved inorganic nitrogen for eight CBNTN stations, comparing the slopes computed using both “heterogeneous” subsamples and “truncated” subsamples with slopes estimated on the basis of “design guideline” subsamples.....	64
6. Ten-year (2003–12) trend slopes for flow-normalized concentration and flux of total phosphorus for eight CBNTN stations, comparing the slopes computed using both “heterogeneous” subsamples and “truncated” subsamples with slopes estimated on the basis of “design guideline” subsamples	65
7. Ten-year (2003–12) trend slopes for flow-normalized concentration and flux of sediment for eight CBNTN stations, comparing the slopes computed using both “heterogeneous” subsamples and “truncated” subsamples with slopes estimated on the basis of “design guideline” subsamples	66

Conversion Factors

SI to Inch/Pound

Multiply	By	To obtain
Length		
kilometer (km)	0.6214	mile (mi)
Area		
square kilometer (km ²)	0.3861	square mile (mi ²)
Volume		
cubic meter (m ³)	35.31	cubic foot (ft ³)
Flow rate		
cubic meter per second (m ³ /s)	35.31	cubic foot per second (ft ³ /s)
Mass		
kilogram (kg)	2.205	pound avoirdupois (lb)
kilogram per year (kg/yr)	2.205	pound per year (lb/yr)
kilogram per year per square kilometer [(kg/yr)/km ²]	5.710	pound per year per square mile [(lb/yr)/mi ²]
kilogram per day per square kilometer [(kg/d)/km ²]	5.710	pound per day per square mile [(lb/d)/mi ²]

Datum

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

Supplemental Information

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

Water year is the period from October 1 to September 30 and is identified by the year in which the period ends. For example, the 2012 water year is October 1, 2011, through September 30, 2012.

Abbreviations

AMLE	Adjusted Maximum Likelihood Estimation
CBNTN	Chesapeake Bay Nontidal Water-Quality Monitoring Network
CBP	Chesapeake Bay Partnership
EPA	U.S. Environmental Protection Agency
NWQL	National Water Quality Laboratory
RIM	River-Input Monitoring
RMSE	root-mean-square error
SSC	suspended-sediment concentration
TMDL	Total Maximum Daily Load
TSS	total suspended solids
USDA	U.S. Department of Agriculture
USGS	U.S. Geological Survey
WRTDS	Weighted Regressions on Time, Discharge, and Season

Acknowledgments

The authors thank the member agencies and jurisdictions of the Chesapeake Bay Partnership for their continued commitment of resources to understanding and protecting the Chesapeake Bay ecosystem.

Jennifer Murphy of the USGS and Qian Zhang of Johns Hopkins University provided thorough technical and editorial reviews, which greatly improved the report. The authors also thank Lori Sprague, Robert Hirsch, and Guoxiang Yang, all of the USGS, for many helpful comments and suggestions. Andy Sekellick of the USGS produced the watershed map.

Application of a Weighted Regression Model for Reporting Nutrient and Sediment Concentrations, Fluxes, and Trends in Concentration and Flux for the Chesapeake Bay Nontidal Water-Quality Monitoring Network, Results Through Water Year 2012

By Jeffrey G. Chanat, Douglas L. Moyer, Joel D. Blomquist, Kenneth E. Hyer, and Michael J. Langland

Abstract

In the Chesapeake Bay watershed, estimated fluxes of nutrients and sediment from the bay's nontidal tributaries into the estuary are the foundation of decision making to meet reductions prescribed by the Chesapeake Bay Total Maximum Daily Load (TMDL) and are often the basis for refining scientific understanding of the watershed-scale processes that influence the delivery of these constituents to the bay. Two regression-based flux and trend estimation models, ESTIMATOR and Weighted Regressions on Time, Discharge, and Season (WRTDS), were compared using data from 80 watersheds in the Chesapeake Bay Nontidal Water-Quality Monitoring Network (CBNTN). The watersheds range in size from 62 to 70,189 square kilometers and record lengths range from 6 to 28 years. ESTIMATOR is a constant-parameter model that estimates trends only in concentration; WRTDS uses variable parameters estimated with weighted regression, and estimates trends in both concentration and flux. WRTDS had greater explanatory power than ESTIMATOR, with the greatest degree of improvement evident for records longer than 25 years (30 stations; improvement in median model R^2 = 0.06 for total nitrogen, 0.08 for total phosphorus, and 0.05 for sediment) and the least degree of improvement for records of less than 10 years, for which the two models performed nearly equally. Flux bias statistics were comparable or lower (more favorable) for WRTDS for any record length; for 30 stations with records longer than 25 years, the greatest degree of improvement was evident for sediment (decrease of 0.17 in median statistic) and total phosphorus (decrease of 0.05). The overall between-station pattern in concentration trend direction and magnitude for all constituents was roughly similar for both models. A detailed case study revealed that trends in concentration estimated by WRTDS can operationally be viewed as a less-constrained equivalent to trends in concentration estimated by ESTIMATOR. Estimates of annual

mean flow-adjusted (ESTIMATOR) and flow-normalized (WRTDS) concentration for years initially constituting the end of a water-quality record showed a similar degree of variability as data for additional years were incrementally added and the initial estimates "aged." On the basis of the results of this broad comparison of the two models, the U.S. Geological Survey is adopting WRTDS as the primary model for estimating constituent fluxes and trends throughout the CBNTN. Nutrient and sediment flux and trend estimates, based on WRTDS, are summarized narratively and tabulated in appendixes for all stations for which fluxes or trends were reported through water year 2012.

WRTDS also was used to explore the sensitivity of flux and trend estimates to three data-quality issues common in many large-scale monitoring networks and evident in some of the CBNTN records. The potential effects of inconsistency in annual sampling effort and inconsistency in storm sampling effort were explored by way of a subsampling experiment using eight of the most densely sampled long-term (1985–2012) stations in the CBNTN as baseline datasets. From each dataset, a set of 10 "design guideline" subsamples was selected, consisting of 12 monthly samples and 8 targeted storm samples per year. The selection was conducted in a manner that preserved the overall intensity of storm sampling in the baseline data. These 10 subsamples were further manipulated to create "heterogeneous" subsamples by removing storm samples prior to 2003. The maximum relative difference between flow-normalized flux estimated in a single year from any of the 10 design guideline subsamples and values estimated in the corresponding year from baseline data was smallest for dissolved inorganic nitrogen (median of 8 stations = 6 percent of baseline estimate), but more appreciable for total phosphorus and sediment (medians of 22 and 32 percent, respectively). The maximum relative difference between flow-normalized flux estimated from the 10 heterogeneous subsamples and values estimated in the

2 Application of a Weighted Regression Model for the Chesapeake Bay Nontidal Water-Quality Monitoring Network

corresponding year from baseline data was more pronounced, with medians for 8 stations of 15, 30, and 53 percent of the corresponding baseline estimates for dissolved inorganic nitrogen, total phosphorus, and sediment, respectively. The worst-case maximum relative differences between flow-normalized flux estimated in a single year from the 10 heterogeneous subsamples and values estimated in the corresponding year from baseline data were 25 percent for dissolved inorganic nitrogen, 37 percent for total phosphorus, and 250 percent for sediment. The results for the heterogeneous subsamples indicate that changes in storm sampling frequency can result in appreciable distortion of estimated trends in flow-normalized flux, especially for total phosphorus and sediment. Trend lines estimated from heterogeneous subsamples tended to converge with the trend lines estimated from baseline data after 2003. In contrast, 2003–12 trends based on subsamples truncated by discarding all data prior to the induced heterogeneity in 2003 showed appreciable biases and differences in slope, relative to the corresponding 2003–12 segment of the trend computed from the design guideline subsamples. Overall, the results indicate that for particulate constituents, load and trend estimates computed using long-term records recently converted to CBNTN design guideline sampling protocols will be most reliable if the trend is computed using the entire record, but reported only for the period that design guideline sampling protocols were followed.

Inconsistencies related to changing laboratory methods were also examined via two manipulative experiments. In the first experiment, increasing and decreasing “stair-step” patterns of changes in censoring level, overall representing a factor-of-five change in the laboratory reporting limit, were artificially imposed on a 27-year record with no censoring and a period-of-record concentration trend of –68.4 percent. Trends estimated on the basis of the manipulated records were broadly similar to the original trend (–63.6 percent for decreasing censoring levels and –70.3 percent for increasing censoring levels), lending a degree of confidence that the survival regression routines upon which WRTDS is based are generally robust to data censoring. The second experiment considered an abrupt disappearance of low-concentration observations of total phosphorus, associated with a laboratory method change and not reflected through censoring, near the middle of a 28-year record. By process of elimination, an upward shift in the estimated flow-normalized concentration trend line around the same time was identified as a likely artifact resulting from the laboratory method change, although a contemporaneous change in watershed processes cannot be ruled out. Decisions as to how to treat records with potential sampling protocol or laboratory methods-related artifacts should be made on a case-by-case basis, and trend results should be appropriately qualified.

Introduction

The Chesapeake Bay is the Nation’s largest estuary, with a watershed area of 166,530 square kilometers, spanning five states. Between 1985 and 2012, population in the watershed

increased from 13.5 to 17.7 million, an increase of about 30 percent, and population is projected to reach 20 million by 2030 (Claggett and Thompson, 2012). Associated with this growth is a list of concerns for the ecological health of the bay and its watershed, including decline in finfish and shellfish populations, loss of vital terrestrial and aquatic habitats, and water-quality degradation, threatening fish and wildlife and creating risks for human health (Chesapeake Bay Program Office, 2014). In response to these concerns, in 2009 the President issued Executive Order 13508 (Obama, 2009) for the protection and restoration of the Chesapeake Bay, effectively putting the full weight of the Federal Government behind the effort to address the environmental challenges facing the Chesapeake Bay. Specifically, the Executive order lists increased delivery of nitrogen, phosphorus, and sediment to the bay from its watershed as the agents most responsible for ecological decline in the estuary. The Chesapeake Bay Total Maximum Daily Load (TMDL; U.S. Environmental Protection Agency, 2010) established by the U.S. Environmental Protection Agency in December 2010 is a keystone of the Federal strategy for action under the Executive order. The TMDL establishes regulatory limits on the delivery of nitrogen, phosphorus, and sediment from member jurisdictions to the bay; these actions are also anticipated to improve the suitability of nontidal waterways for aquatic biota, recreation, and public water supply.

Through its partnership with other Federal agencies, bay State and local agencies, non-profit organizations, and academic institutions in the Chesapeake Bay Partnership (CBP), the U.S. Geological Survey (USGS) plays a key role in strengthening scientific support for decision making to restore the bay and its watershed. Among its fundamental roles is to work closely with the U.S. Environmental Protection Agency (EPA) and U.S. Department of Agriculture (USDA) to “assess and explain water-quality conditions and change” (U.S. Geological Survey, 2010). To fulfill this role, the USGS Chesapeake Study Plan for fiscal years 2011–16 specifies that the agency focus its scientific resources on (1) enhancing models of nutrients, sediment, and groundwater to help target management actions, (2) improving regional water-quality monitoring to assess progress toward water-quality goals, (3) establishing monitoring in small urban and agricultural watersheds to assess the effects of restoration activities, and (4) explaining water-quality changes and the effects of basin-wide TMDL implementation (U.S. Geological Survey, 2010). Scientific studies and stakeholder engagement relevant to the regional and basinwide components of the USGS Chesapeake Study Plan are supported primarily by raw data, analysis, and communication products provided by the Chesapeake Bay Nontidal Water-Quality Monitoring Network (CBNTN; fig. 1). The CBNTN is a collaborative effort through which water-quality data collected by the USGS, bay States, District of Columbia, and the Susquehanna River Basin Commission are assembled for the purpose of estimating constituent fluxes (mass passing a gaged nontidal location per unit time; also termed loads), as well as trends in fluxes over time, throughout

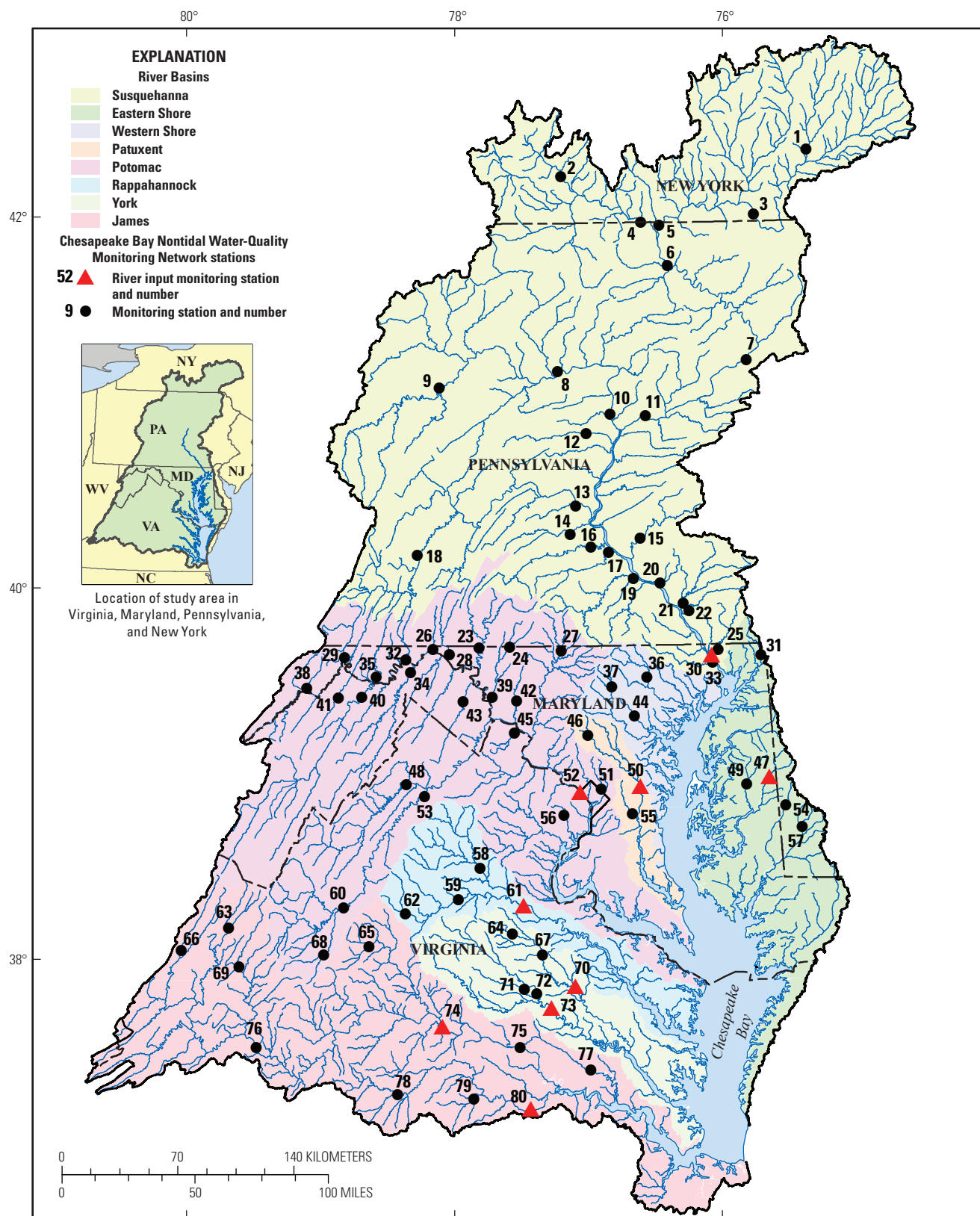


Figure 1. Locations of major drainage basins in the Chesapeake Bay watershed and locations of Chesapeake Bay Nontidal Water-Quality Monitoring Network (CBNTN) stations for which flux and (or) trend estimates were published through water year 2012.

4 Application of a Weighted Regression Model for the Chesapeake Bay Nontidal Water-Quality Monitoring Network

the bay watershed using consistent and defensible methods. The history of the CBNTN dates to the mid-1980s, when the USGS began reporting fluxes and trends at nine nontidal stations near the Fall Line in Maryland and Virginia. These stations, known collectively as the “River Input Monitoring” (RIM) network, represent about 78 percent of the streamflow entering the bay from the nontidal portion of the watershed (Langland and others, 1995). In 2004, the CBNTN was formalized with the addition of 21 upland stations with long-term data, specification of required sampling and analysis protocols, and identification of additional stations suitable for future network expansion with a long-term vision of a total of 188 stations (Nontidal Water Quality Monitoring Workgroup, written commun., September, 2004). Growth of the network since its formalization has proceeded largely in accordance with the 2004 plan with some modification in station selection to meet the evolving scientific needs of the CBP (Chesapeake Bay Program Office, 2009). Current network design guidelines require that a station have a water-quality record of at least 5 years to estimate flux and 10 years to estimate a trend. In water year 2012, flux and (or) trend estimates for nutrients and sediment were made for a total of 80 stations, with the youngest stations having data beginning water year 2007 (fig. 1; table 1). An additional 36 stations have been added to the network since water year 2007, but these do not yet meet the 5-year minimum requirement for flux estimation (table 2); this brings the total number of active CBNTN stations to 116 as of fall 2014. Under the current data requirements for reporting fluxes and trends, flux estimates through water year 2014 will be available for 82 stations, and trend estimates through water year 2014 will be available for 58 stations (fig. 2).

Constituent fluxes at CBNTN stations are estimated on a daily time scale using regression techniques (Helsel and Hirsch, 2002). In brief, an observed relation between sampled constituent concentration and daily discharge based on sparse sampling (typically about 20 sampled days per year) is used to estimate concentration, conditioned on daily discharge, on unsampled days. Daily estimates of concentration and flux, the product of concentration and discharge, are then aggregated by averaging (for concentration) or summation (for flux), typically at monthly or annual time scales. Historically, annual and monthly fluxes at CBNTN stations have been estimated using ESTIMATOR, a 7-parameter regression equation originally described by Cohn and others (1992). Regression parameters in ESTIMATOR include an intercept, linear and quadratic terms for both time and discharge, and sine and cosine terms representing seasonal variability. Coefficients on these terms are constant; that is, a single set of parameters is fit to the entire observed water-quality record. The model reports monthly and annual flux estimates with uncertainty bounds, which are derived from the uncertainty in daily flux estimates by using assumptions about the correlation structure of daily fluxes (Gilroy and others, 1990).

More recently, Hirsch and others (2010) published an alternate flux and trend estimation model, Weighted Regressions on Time, Discharge, and Season (WRTDS). The WRTDS

model was developed to solve a set of generic problems with statistical models used for long-term water-quality trend estimation. Initially, the model was applied to Chesapeake Bay RIM datasets to demonstrate the model’s usefulness in describing a variety of features of water-quality records. Like ESTIMATOR, the model uses a regression approach, but WRTDS allows for time-variable regression parameters, tailored for the particular day of the record for which a concentration estimate is required. Each set of parameters is estimated using a subset of the water-quality record, with observations weighted on the basis of their “similarity,” in terms of time, discharge, and season, to the conditions on the day for which the estimate is required. Partly on the basis of this flexibility and relative sophistication, WRTDS has rapidly gained acceptance in the research community as a tool for graphical data analysis and flux/trend estimation in support of studies of watershed-scale transport of nutrients, sediment, and other contaminants (Sprague and others, 2011; Zhang and others, 2013; Green and others, 2014; Murphy and others, 2014; Corsi and others, 2015). WRTDS was originally conceptualized as an exploratory tool, and the original model formulation did not provide estimates of statistical uncertainty in flux or trend results; this limitation is currently the topic of active research and development.

Partly because the model was developed and originally presented using long-term datasets from the Chesapeake Bay RIM stations, WRTDS came under consideration soon after its publication as a potential replacement for ESTIMATOR throughout the CBNTN. Moyer and others (2012) directly compared ESTIMATOR and WRTDS in terms of explanatory power (the proportion of the variability of the concentration observations about their mean that is explained by the model) and flux bias (the net tendency of a model to over- or underpredict the observed fluxes). Focusing primarily on estimated fluxes, Moyer and others (2012) showed that, overall, WRTDS had both appreciably greater explanatory power and appreciably lower bias than ESTIMATOR given the same water-quality record. Considering the problem of flux bias in particular, Hirsch (2014) found that in many cases the two methods had similar amounts of bias, but in some cases WRTDS had substantially lower bias than ESTIMATOR. On the basis of these two studies, the USGS adopted WRTDS as the primary tool for estimating fluxes across the CBNTN; WRTDS flux estimates based on data collected through water year 2012 were first published online in December 2013 (Langland and others, 2013).

Estimation of trends in concentration or flux over time is somewhat more complicated than estimation of either quantity at a fixed point in time. Because the regression methods are based primarily on the relation between concentration and discharge, credible trend estimation is complicated by the need to minimize the confounding effect of any corresponding trend in discharge. For concentration trends in ESTIMATOR, this is accomplished through direct interpretation of the coefficients on the linear and quadratic time terms in the regression model; uncertainty bounds around the resulting trend estimate

Table 1. Names, physical characteristics, period of record, and model selection for Chesapeake Bay Nontidal Water-Quality Monitoring Network stations for which flux and (or) trend estimates were published through water year 2012.

[Stations are listed in downstream order by U.S. Geological Survey station identification number. Latitude and longitude coordinates, in decimal degrees, are referenced to the North American Datum of 1983. Starting water year: Blank cell indicates constituent was not modeled at that station; *, Model was disqualified for lack-of-fit. Sediment data type: TSS, total suspended solids; SSC, suspended-sediment concentration. Model(s) used (refer to section “Description of Scenarios Considered”): E, ESTIMATOR; W, WRTDS [Weighted Regressions on Time, Discharge, and Season]; Flux: Multiple entries indicate that model choice was constituent specific. Entries are blank for *10_YEAR_SECONDARY* stations, for which no flux results were reported; Trend: Blank cells indicate no trend results were reported because the water-quality record length was less than 10 years. Pink shading denotes River Input Monitoring (RIM) station; gray shading denotes *10_YEAR_SECONDARY* station]

Station ID	Station name	Map key (fig. 1)	Abbreviation	Latitude (decimal degrees)	Longitude (decimal degrees)	Drainage area (square kilometers)	Starting water year			Sediment data type	Model(s) used	
							Total nitrogen	Total phosphorus	Sediment		Flux	Trend
01487000	Nanticoke River near Bridgeville, DE	57	NANT	38.7283	-75.5619	195	2006	2006	2006	TSS	E	
01488500	Marshyhope Creek near Adamsville, DE	54	MARS	38.8497	-75.6731	121	2006	2006	2006	TSS	E	
01491000	Choptank River near Greensboro, MD	47	CHOP	38.9972	-75.7858	293	1985	1985	1985	SSC	W	E
01491500	Tuckahoe Creek near Ruthsburg, MD	49	TUCK	38.9668	-75.9431	221	2006	2006	2006	TSS	W	
01495000	Big Elk Creek at Elk Mills, MD	31	BGEL	39.6571	-75.8224	136	2006	2006	2006	TSS	W	
01502500	Unadilla River at Rockdale, NY	1	UNAD	42.3778	-75.4061	1,347	2006	2006	2006	TSS	E	
01503000	Susquehanna River at Conklin, NY	3	SRCK	42.0353	-75.8031	5,781	2006	2006	2006	TSS	W,E	
01515000	Susquehanna River near Waverly, NY	5	SRWA	41.9847	-76.5011	12,362	2005	2005	2005	SSC	W,E	
01529500	Cohocton River near Campbell, NY	2	COHO	42.2525	-77.2167	1,217	2006	2006	2006	TSS	E	
01531000	Chemung River at Chemung, NY	4	CHEM	42.0022	-76.6347	6,491	2005	2005	2005	SSC	W,E	
01531500	Susquehanna River at Towanda, PA	6	SRTW	41.7654	-76.4408	20,194	1985	1985	1989	SSC	W	E
01536500	Susquehanna River at Wilkes-Barre, PA	7	SRWB	41.2509	-75.8808	25,796	1989	1985	2004	SSC	W	E
01540500	Susquehanna River at Danville, PA	11	SRDA	40.9581	-76.6191	29,060	1985	1985	1985	SSC	W	E
01542500	West Branch Susquehanna River at Karthaus, PA	9	WBSK	41.1176	-78.1089	3,787	2005	2005	2005	SSC	W	
01549760	West Branch Susquehanna River near Jersey Shore, PA	8	WBSJ	41.2023	-77.2511	13,533	2006	2006	2006	SSC	E	
01553500	West Branch Susquehanna River at Lewisburg, PA	10	WBSL	40.9676	-76.8764	17,734	1985	1985	1985	SSC	W	E
01555000	Penns Creek at Penns Creek, PA	12	PENN	40.8667	-77.0483	780	2005	2005	2005	SSC	W	
01562000	Raystown Branch Juniata River at Saxton, PA	18	RBJU	40.2159	-78.2653	1,958	2005	2005	2005	SSC	W	
01567000	Juniata River at Newport, PA	13	JUNI	40.4784	-77.1291	8,687	1985	1985	1985	SSC	W	E
01568000	Sherman Creek at Shermans Dale, PA	14	SHER	40.3234	-77.1689	536	2005	2005	2005	SSC	W	
01570000	Conodoguinet Creek near Hogestown, PA	16	COND	40.2522	-77.0211	1,217	2005	2005	2005	SSC	W	
01571500	Yellow Breeches Creek near Camp Hill, PA	17	YELL	40.2248	-76.8980	552	2005	2005	2005	SSC	W	
01573560	Swatara Creek near Hershey, PA	15	SWAT	40.2984	-76.6677	1,251	2005	2005	2005	SSC	W	
01574000	West Conewago Creek near Manchester, PA	19	WCWG	40.0823	-76.7199	1,321	2005	2005	2005	SSC	W	
01576000	Susquehanna River at Marietta, PA	20	SRMA	40.0545	-76.5308	67,314	1987	1987	1987	SSC	W	E
01576754	Conestoga River at Conestoga, PA	21	CONE	39.9465	-76.3677	1,217	1985	1985	1985	SSC	W	E
01576787	Pequea Creek near Martic Forge, PA	22	PEQU	39.9059	-76.3283	383	2005	2005	2005	SSC	W	
01578310	Susquehanna River at Conowingo, MD	30	SRCW	39.6579	-76.1742	70,189	1985	1985	1985	SSC	W	E
01578475	Octoraro Creek at Richardsmere, MD	25	OCTO	39.7068	-76.1152	458	2007	2007	2007	TSS	E	

6 Application of a Weighted Regression Model for the Chesapeake Bay Nontidal Water-Quality Monitoring Network

Table 1. Names, physical characteristics, period of record, and model selection for Chesapeake Bay Nontidal Water-Quality Monitoring Network stations for which flux and (or) trend estimates were published through water year 2012.—Continued

[Stations are listed in downstream order by U.S. Geological Survey station identification number. Latitude and longitude coordinates, in decimal degrees, are referenced to the North American Datum of 1983. Starting water year: Blank cell indicates constituent was not modeled at that station; *, Model was disqualified for lack-of-fit. Sediment data type: TSS, total suspended solids; SSC, suspended-sediment concentration. Model(s) used (refer to section “Description of Scenarios Considered”): E, ESTIMATOR; W, WRTDS [Weighted Regressions on Time, Discharge, and Season]; Flux: Multiple entries indicate that model choice was constituent specific. Entries are blank for *10_YEAR_SECONDARY* stations, for which no flux results were reported; Trend: Blank cells indicate no trend results were reported because the water-quality record length was less than 10 years. Pink shading denotes River Input Monitoring (RIM) station; gray shading denotes *10_YEAR_SECONDARY* station]

Station ID	Station name	Map key (fig. 1)	Abbreviation	Latitude (decimal degrees)	Longitude (decimal degrees)	Drainage area (square kilometers)	Starting water year			Sediment data type	Model(s) used	
							Total nitrogen	Total phosphorus	Sediment		Flux	Trend
01580520	Deer Creek near Darlington, MD	33	DEER	39.6174	−76.1919	425	2006	2006	2006	TSS	W	
01582500	Gunpowder Falls at Glencoe, MD	36	GUNP	39.5497	−76.6361	414	1985	1985	1985	TSS	W	E
01586000	North Branch Patapsco River at Cedarhurst, MD	37	NBPA	39.5037	−76.8849	147	1985	1985	1985	TSS	W	E
01589300	Gwynns Falls at Villa Nova, MD	44	GWYN	39.3459	−76.7332	84	2003	2003	2003	TSS	W	E
01591000	Patuxent River near Unity, MD	46	PAXU	39.2383	−77.0557	90	1985	1985	1985	TSS	W	E
01594440	Patuxent River near Bowie, MD	50	PAXB	38.9559	−76.6937	901	1985	1985	1985	SSC	W	E
01594526	Western Branch at Upper Marlboro, MD	55	WBUM	38.8142	−76.7487	232	2006	2006	2006	TSS	W	
01599000	Georges Creek at Franklin, MD	38	GEOR	39.4939	−79.0447	188	1985	1985	1985	TSS	W	E
01601500	Wills Creek near Cumberland, MD	29	WILL	39.6696	−78.7880	640	1985	1985	1985	TSS	W	E
01604500	Patterson Creek near Headsville, WV	41	PATT	39.4431	−78.8219	572	2006	2006	2006	SSC	W	
01608500	South Branch Potomac River near Springfield, WV	40	SBPO	39.4470	−78.6542	3,784	2006	2006	2006	SSC	W,E	
01609000	Town Creek near Oldtown, MD	35	TOWN	39.5532	−78.555	383	2007	2007	2007	SSC	E	
01610155	Sideling Hill Creek near Bellegrove, MD	32	SIDE	39.6495	−78.3441	264	2007	2007	2007	TSS	E	
01611500	Cacapon River near Great Cacapon, WV	34	CACA	39.5823	−78.3097	1,748	2006	2006	2006	SSC	W,E	
01613095	Tonoloway Creek near Hancock, MD	26	TONO	39.7064	−78.1528	287	2006	2006	2006	TSS	E	
01613525	Licking Creek near Pectonville, MD	28	LICK	39.6763	−78.0425	500	2006	2006	2006	TSS	W	
01614500	Conococheague Creek at Fairview, MD	23	CONC	39.7164	−77.8248	1,279	1985	1985	1985	TSS	W	E
01616500	Opequon Creek near Martinsburg, WV	43	OPEQ	39.4237	−77.9386	707	2006	2006	2006	SSC	W,E	
01619000	Antietam Creek near Waynesboro, PA	24	ANTW	39.7163	−77.6066	242	2006	2006	2006	TSS	E	
01619500	Antietam Creek near Sharpsburg, MD	39	ANTS	39.4498	−77.7302	728	1985	1985	1985	TSS	W	E
01626000	South River near Waynesboro, VA	68	SOUT	38.0576	−78.9081	329	2003	2003	2003	TSS		E
01628500	South Fork Shenandoah River near Lynnwood, VA	60	SFSL	38.3226	−78.7547	2,795	2003	2003	2003	TSS		E
01631000	South Fork Shenandoah River at Front Royal, VA	53	SFSF	38.9140	−78.2108	4,232	1985	1996	2003	SSC	W	E
01634000	North Fork Shenandoah River near Strasburg, VA	48	NFSH	38.9768	−78.3361	1,994	1985	1996	2003	SSC	W	E
01637500	Catoctin Creek near Middletown, MD	42	CATM	39.4273	−77.5562	173	1985	1985	1985	TSS	W	E
01638480	Catoctin Creek near Taylorstown, VA	45	CATT	39.2551	−77.5764	232	2003	2003	2003	TSS		E
01639000	Monocacy River at Bridgeport, MD	27	MONO	39.6788	−77.2345	448	1985	1985	1985	TSS	W	E
01646580	Potomac River at Chain Bridge, at Washington, DC	52	POTC	38.9296	−77.1169	29,966	1985	1985	1985	SSC	W	E

Table 1. Names, physical characteristics, period of record, and model selection for Chesapeake Bay Nontidal Water-Quality Monitoring Network stations for which flux and (or) trend estimates were published through water year 2012.—Continued

[Stations are listed in downstream order by U.S. Geological Survey station identification number. Latitude and longitude coordinates, in decimal degrees, are referenced to the North American Datum of 1983. Starting water year: Blank cell indicates constituent was not modeled at that station; *, Model was disqualified for lack-of-fit. Sediment data type: TSS, total suspended solids; SSC, suspended-sediment concentration. Model(s) used (refer to section “Description of Scenarios Considered”): E, ESTIMATOR; W, WRTDS [Weighted Regressions on Time, Discharge, and Season]; Flux: Multiple entries indicate that model choice was constituent specific. Entries are blank for *10_YEAR_SECONDARY* stations, for which no flux results were reported; Trend: Blank cells indicate no trend results were reported because the water-quality record length was less than 10 years. Pink shading denotes River Input Monitoring (RIM) station; gray shading denotes *10_YEAR_SECONDARY* station]

Station ID	Station name	Map key (fig. 1)	Abbreviation	Latitude (decimal degrees)	Longitude (decimal degrees)	Drainage area (square kilometers)	Starting water year			Sediment data type	Model(s) used	
							Total nitrogen	Total phosphorus	Sediment		Flux	Trend
01651000	Northwest Branch Anacostia River near Hyattsville, MD	51	NWBA	38.9523	−76.9661	128	2007	2007	2007	SSC	W	
01654000	Accotink Creek near Annandale, VA	56	ACCO	38.8129	−77.2283	62	2003	2003	2003	TSS		E
01664000	Rappahannock River at Remington, VA	58	RPHR	38.5307	−77.8136	1,603	2003	2003	2003	TSS		E
01665500	Rapidan River near Ruckersville, VA	62	RPDR	38.2807	−78.3400	298	2003	2003				E
01667500	Rapidan River near Culpeper, VA	59	RPDC	38.3504	−77.9749	1,212	2005	2005	2005	SSC	W	
01668000	Rappahannock River near Fredericksburg, VA	61	RPHF	38.3085	−77.5292	4,131	1985	1985	2003	SSC	W	E
01671020	North Anna River at Hart Corner near Doswell, VA	72	NANN	37.8501	−77.4278	1,197	1985	1994	1985	TSS	W	E
01671100	Little River near Doswell, VA	71	LITT	37.8726	−77.5130	277	2003	2003	2003	TSS		E
01673000	Pamunkey River near Hanover, VA	73	PAMU	37.7676	−77.3322	2,792	1985	1985	2003	SSC	W	E
01673800	Po River near Spotsylvania, VA	64	PORV	38.1715	−77.5947	201	2003	2003				E
01674000	Mattaponi River near Bowling Green, VA	67	MABG	38.0618	−77.3858	663	2003	2003				E
01674500	Mattaponi River near Beulahville, VA	70	MABU	37.8839	−77.1653	1,562	1985	1985	2003	SSC	W	E
02011500	Back Creek near Mountain Grove, VA	66	BACK	38.0696	−79.8970	347	2003					E
02015700	Bullpasture River at Williamsville, VA	63	BULL	38.1954	−79.5703	285	2003	2003*				E
02020500	Calpasture River above Mill Creek At Goshen, VA	69	CALF	37.9879	−79.4937	365	2003	2003*				E
02024752	James River at Blue Ridge Parkway near Big Island, VA	76	JAMB	37.5554	−79.3673	7,967	2006	2006	2006	SSC	W	
02031000	Mechums River near White Hall, VA	65	MECH	38.1026	−78.5928	247	2003	2003	2003	TSS		E
02035000	James River at Cartersville, VA	74	JAMC	37.6709	−78.0858	16,193	1985	1985	2003	SSC	W	E
02037500	James River near Richmond, VA	75	JAMR	37.5632	−77.5469	17,490	1985	1994	1985	TSS	W	E
02039500	Appomattox River at Farmville, VA	78	APPF	37.3071	−78.3886	782	2003	2003	2003	TSS		E
02041000	Deep Creek near Mannboro, VA	79	DEEP	37.2832	−77.8697	409	2003	2003	2003	TSS		E
02041650	Appomattox River at Matoaca, VA	80	APPM	37.2252	−77.4753	3,476	1985	1985	2003	SSC	W	E
02042500	Chickahominy River near Providence Forge, VA	77	CHIC	37.4363	−77.0608	650	1985	1985	1985	TSS	W	E

8 Application of a Weighted Regression Model for the Chesapeake Bay Nontidal Water-Quality Monitoring Network

Table 2. Names, physical characteristics, and period of record for Chesapeake Bay Nontidal Water-Quality Monitoring Network stations active as of fall 2014.

[Latitude and longitude coordinates, in decimal degrees, are referenced to the North American Datum of 1983]

Station ID	Station name	Latitude (decimal degrees)	Longitude (decimal degrees)	Drainage area (square kilometers)	Starting water year		
					Total nitrogen	Total phosphorus	Sediment
01486000	Manokin Branch near Princess Anne, MD	38.2139	-75.6714	12	2011	2011	2011
01493112	Chesterville Branch near Crumpton, MD	39.6986	-77.3300	16	2011	2011	2011
01493500	Morgan Creek near Kennedyville, MD	39.2800	-76.0146	33	2011	2011	2011
01511500	Tioughnioga River at Itaska, NY	42.2981	-75.9092	1,891	2012	2012	2012
01553700	Chillisquaque Creek at Washingtonville, PA	40.9744	-76.8000	133	2012	2012	2012
01555500	East Mahantango Creek near Dalmatia, PA	40.6111	-76.9122	420	2012	2012	2012
01565000	Kishacoquillas Creek at Reedsville, PA	40.6547	-77.5833	425	2012	2012	2012
01571000	Paxton Creek near Penbrook, PA	40.3083	-76.8500	29	2012	2012	2012
01573160	Quittapahilla Creek near Bellegrove, PA	41.3553	-76.0142	192	2012	2012	2012
01573695	Conewago Creek near Bellaire, PA	40.1953	-76.5678	53	2011	2011	2011
01573710	Conewago Creek near Falmouth, PA	40.1511	-76.6894	123	2011	2011	2011
01575585	Codorus Creek at Pleasureville, PA	40.0186	-76.6733	692	2012	2012	2012
015765195	Big Spring Run near Mylin Corners, PA	39.9959	-76.2640	4	2011	2011	2011
01576787	Pequea Creek at Martie Forge, PA	41.6289	-77.6561	181	2012	2012	2012
01577500	Muddy Creek at Castle Fin, PA	39.7725	-76.3161	344	2014	2014	2014
01581752	Plumtree Run near Bel Air, MD	39.4964	-76.3478	6	2012	2012	2012
0158175320	Wheel Creek near Abingdon, MD	39.4817	-76.3405	2	2011	2011	2011
01593500	Little Patuxent River at Guilford, MD	39.1678	-76.8513	98	2011	2011	2011
01595300	Abram Creek at Oakmont, WV	39.3667	-76.1542	110	2012	2012	2012
01601100	Wills Creek at Eilerslie, MD	39.7185	-78.7710	479	2011	2011	2011
01613030	Warm Springs Run near Berkeley Springs, WV	39.6406	-78.2189	18	2011	2011	2011
01614000	Back Creek near Jones Springs, WV	39.5119	-78.0375	609	2012	2012	2012
01616400	Mill Creek at Bunker Hill, WV	39.3346	-78.0534	48	2011	2011	2011
01618100	Rockymarsh Run at Scrabble, WV	39.4831	-77.8318	41	2011	2011	2011
01621050	Muddy Creek at Mount Clinton, VA	38.4867	-78.9606	37	2011	2011	2011
01632900	Smith Creek near New Market, VA	38.6935	-78.6428	242	2010	2010	2010
01636500	Shenandoah River at Millville, WV	39.2819	-77.7894	7,876	2013	2014	2015
01646000	Difficult Run near Great Falls, VA	38.9758	-77.2461	150	2011	2011	2011
01648000	Rock Creek at Sherill Drive at Washington, DC	38.9725	-77.0400	161	2013	2013	2013
01651770	Hickey Run at National Arboretum at Washington, DC	38.9171	-76.9693	3	2013	2013	2013
01651800	Watts Branch at Washington, DC	38.9011	-76.9422	8	2013	2013	2013
01658000	Mattawoman Creek near Pomomkey, MD	38.5961	-77.0560	142	2010	2010	2010
01658500	South Fork Quantico Creek near Independent Hill, VA	38.5872	-77.4289	20	2011	2011	2011
01669520	Dragon Swamp at Mascot, VA	37.6336	-76.6967	282	2011	2011	2011
01674182	Polecat Creek at Route 301 near Penola, VA	37.9603	-77.3436	127	2012	2012	2012
02034000	Rivanna River at Palmyra, VA	37.8579	-78.2658	1,717	2011	2012	2013

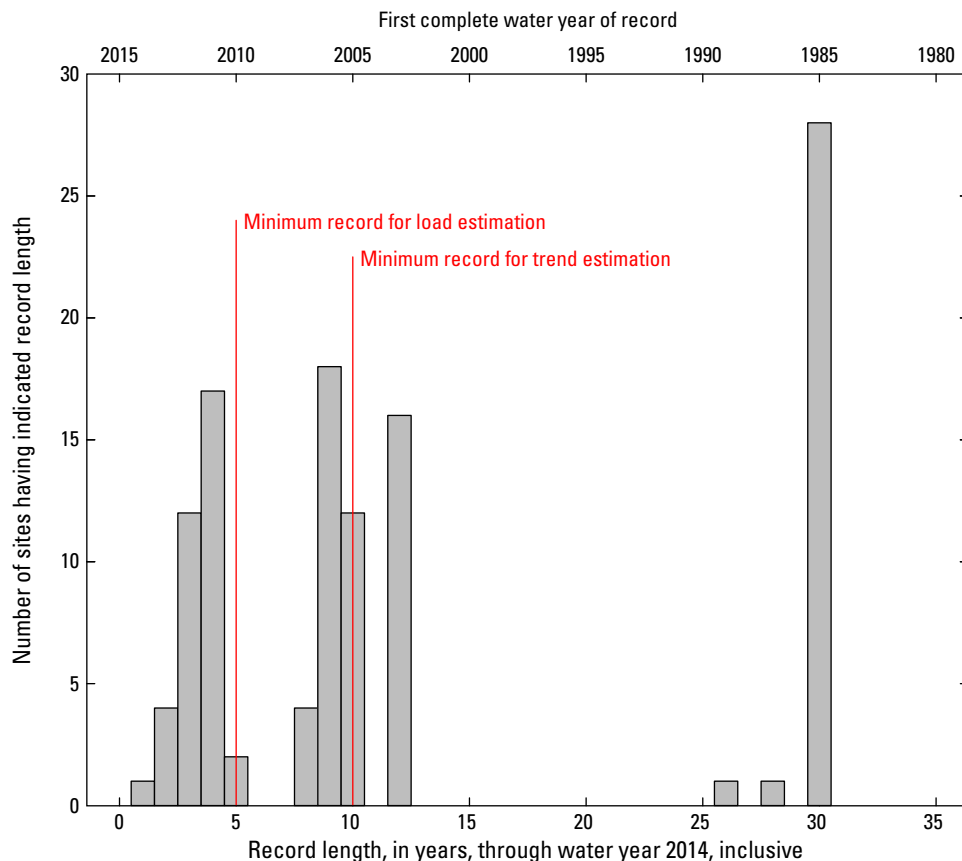


Figure 2. Water-quality record lengths for all Chesapeake Bay Nontidal Water-Quality Monitoring Network (CBNTN) stations active as of fall 2014.

can also be analytically derived from the uncertainty in the estimates for these coefficients (Langland and others, 2006). However, the functional form of the regression model constrains estimated trends in flux to be the same as estimated trends in concentration (when both are expressed in percentage change per year); in reality these trends can differ substantially. These differences can reveal environmentally relevant processes that are underway such as changes in point-source contributions, changes in groundwater contributions, or changes in high-flow-related nonpoint-source contributions. In WRTDS, variation of the regression coefficients with time, discharge, and season necessitates a more complex algorithm for trend estimation. This added complexity, however, conveys several potential advantages: First, the time trend is not constrained to follow a linear/parabolic trajectory, as is the case with ESTIMATOR. Second, WRTDS can provide independent estimates of trend in both concentration and flux. Despite these advantages, the novelty of the trend estimation algorithm in WRTDS has thus far hindered direct comparison with the more straightforward method used in ESTIMATOR, and no published studies comparing the two models' estimated trends, analogous to recent comparisons of estimated flux made by Moyer and others (2012) and Hirsch (2014), exist. Further,

confidence bounds on trends, available using ESTIMATOR, are still under development for WRTDS. For these reasons, and because CBNTN trend products are anticipated to receive increasing visibility in the water-resources management community in light of potential regulatory action under the bay TMDL, the USGS decided to retain the more-established ESTIMATOR model for CBNTN trend reporting through water year 2012.

Purpose and Scope

The purpose of this report has two main areas of focus. The first is to compare the performance of WRTDS and ESTIMATOR across the CBNTN, with particular attention to trends. The structures of the two models are examined in detail, and insights gained in the study are used to better understand between-model differences in estimated fluxes and trends. On the basis of that comparison, WRTDS was chosen as a replacement for ESTIMATOR as the primary tool for future reporting of both flux and trend results for the CBNTN, and flux and trend results computed exclusively with WRTDS for all CBNTN stations through water year 2012 are provided. The second main focus of this report is to explore and

document some of the potential effects of inconsistencies in sampling and laboratory analysis protocols—features common in any large operational water-quality network—on modeled flux and trend results. Records with dense sub-daily sampling, variability in annual sampling effort, and variability in storm sampling effort were considered separately, and the effects of some common instances of changes in laboratory methods were explored. Lessons learned through those explorations can be used as the basis for future CBNTN data analysis and reporting.

The objectives of this report are to

1. Document the technical considerations leading to the adoption of the WRTDS model as the primary tool for reporting both fluxes and trends for nutrients and sediment at the CBNTN stations;
2. Report the WRTDS estimates of water year 2012 fluxes, along with the long-term and 10-year trends in concentration and flux for the CBNTN stations, for use as the primary basis for both management decision support and interpretive studies exploring factors underlying observed trends; and
3. Identify technical issues associated with using WRTDS across the CBNTN, quantify the potential impact of these issues, and suggest potential solutions.

This report is intended primarily for researchers, watershed modelers, and land managers intent upon interpreting nutrient and sediment flux, yield, and trend results generated by regression-based methods, in particular by WRTDS. The report presumes some familiarity with the basic theory of regression-based approaches for flux and trend estimation (for example, Cohn and others, 1989; Helsel and Hirsch, 2002).

Technical Considerations Leading to the Adoption of WRTDS

In this section of the report, the performance of WRTDS and ESTIMATOR are compared across the CBNTN. First, each model's structure and implementation are reviewed. The designs of the comparative flux estimation studies of Moyer and others (2012) and Hirsch (2014), along with their major conclusions are then summarized. Using datasets through water year 2012 from across the CBNTN, the advantages of using WRTDS as a flux estimation tool, reported by Moyer and others (2012) and Hirsch (2014) using carefully selected water-quality records, are shown to apply when WRTDS is applied across a diverse operational water-quality network. Trends are then considered, first with a comparison of the CBNTN trends in concentration through water year 2012, reported previously using ESTIMATOR, with the corresponding concentration trends computed using WRTDS. The differences in trend estimation approaches between the two models

are reconciled with a detailed case study, and the factors that limit ESTIMATOR to reporting trends in concentration, but not trends in flux, are discussed. Trends reported by both WRTDS and ESTIMATOR are sensitive, through different mechanisms but to a similar degree, to incremental addition of new data to an existing long-term record. Finally, on the basis of all these findings, WRTDS is chosen to supersede ESTIMATOR as the primary tool for future reporting of both flux and trend results for the CBNTN effective on the next reporting cycle, currently projected to include data through water year 2014. Annual flux and trend estimates made using WRTDS for all CBNTN stations through water year 2012 are provided as an appendix to this report, with the intent that they become the results of record, for both assessment of progress toward water-quality goals and for scientific interpretation, in the interim.

Overview of ESTIMATOR and WRTDS Models

Cohn and others (1992) presented a seven-parameter log-linear model for daily water-quality constituent concentrations, which served as the foundation of the FORTRAN program ESTIMATOR, used throughout the CBNTN. (ESTIMATOR is mathematically equivalent to the seven-parameter version of LOADEST; Runkel and others, 2004.) The ESTIMATOR model has the form

$$\ln(c) = \hat{\beta}_0 + \hat{\beta}_1 \ln(q / q_c) + \hat{\beta}_2 [\ln(q / q_c)]^2 + \hat{\beta}_3(t - t_c) + \hat{\beta}_4(t - t_c)^2 + \hat{\beta}_5 \sin(2\pi t) + \hat{\beta}_6 \cos(2\pi t) + \varepsilon \quad (1)$$

where

\ln	is the natural log function;
c	is the measured concentration, in milligrams per liter;
q	is measured daily-mean discharge, in cubic feet per second;
t	is time, in decimal years;
q_c, t_c	are centering variables for streamflow and time;
$\hat{\beta}$	are coefficients estimated by ordinary least squares (uncensored observation) and Adjusted Maximum Likelihood Estimation (AMLE) (censored observations);
$\hat{\beta}_0$	is a constant;
$\hat{\beta}_1, \hat{\beta}_2$	describe the relation between concentration and streamflow;
$\hat{\beta}_3, \hat{\beta}_4$	describe the relation between concentration and time, independent of flow;
$\hat{\beta}_5, \hat{\beta}_6$	describe seasonal variation in concentration; and
ε	is the unexplained variation.

Cohn and others (1992) demonstrated the effectiveness of this model for estimating nitrogen and phosphorus at several Chesapeake Bay RIM stations. Variation in constituent concentration as a function of discharge is addressed by including both discharge and discharge-squared terms, thus allowing the model to represent linear or parabolic relations between discharge and concentration. Variation in concentration as a function of time is accounted for by including both time and time-squared terms; these terms allow the model to represent linear or parabolic trends over time. Seasonal variation is represented using one sine and one cosine term, allowing for the representation of a single sinusoidal trend with fixed amplitude and fixed phase. ESTIMATOR reports concentration and mass flux on a daily basis, with uncertainty bounds derived from the regression model standard error. Aggregated estimates at monthly or annual scales are also provided with uncertainty bounds, using additional assumptions about model uncertainty, parameter uncertainty, and the serial correlation structure of daily estimates (Gilroy and others, 1990). Potential bias associated with transformation of predicted concentrations from logarithmic to arithmetic space is addressed using a single multiplicative correction, which is based on the standard error of the model residuals (Cohn and others, 1992). As is the case with nearly all regression-based flux and trend estimation models, a key statistical requirement for the valid application of ESTIMATOR is that water-quality observations be collected at discharges that are as representative as possible of the entire range of daily discharges encountered over the period of the simulation; to ensure this requirement, sampling protocols are often stratified in favor of infrequent high flows (Langland and others, 2006).

It has been well established that the concentration of many water-quality constituents is a function of discharge and that identification of a time trend in constituent concentration benefits from an approach which accounts for the variability in discharge (Helsel and Hirsch, 2002) particularly when there are multiyear periods of relatively high- or low-flow conditions. In ESTIMATOR, this is accomplished by direct interpretation of the coefficients on the time and time-squared terms ($\hat{\beta}_3$ and $\hat{\beta}_4$) in equation 1; the resulting trend is called the “trend in flow-adjusted concentration” and is reported as a percentage change relative to a concentration on a specified starting date. As all the model coefficients in equation 1 are constants determined by best fit to the data, and the only other explanatory variables in equation 1 are discharge and season, a flow-adjusted trend in concentration could be defined simply as “the trend in the observed concentration data that cannot be explained by discharge or season.” Langland and others (2006) demonstrate the analytic calculation of a flow-adjusted trend beginning at time t_0 and ending at t , from equation 1, as

$$FAT_{t_0,t} = 100 \cdot (e^{\hat{\tau}_{FA}} - 1) \quad (2)$$

where

$$\hat{\tau}_{FA} = \hat{\beta}_3(t - t_0) + \hat{\beta}_4[(t - t_c)^2 - (t_0 - t_c)^2] \quad (3)$$

where $\hat{\beta}_3$ and $\hat{\beta}_4$ are the fitted coefficients, t_c is the time centering constant, and $\hat{\tau}_{FA}$ is the estimated change in the log of concentration between time t_0 and t . The calculation includes a means of propagating estimation uncertainty in $\hat{\beta}_3$ and $\hat{\beta}_4$ into uncertainty bounds, also expressed as percentage changes, around the flow-adjusted trend. Significantly, $\hat{\tau}_{FA}$ has the effect of shifting the concentration at all discharges up or down equally; that is, it does not influence the parabolic shape of the concentration-discharge relation, which is fixed by parameters $\hat{\beta}_1$ and $\hat{\beta}_2$ (eq. 1). Moreover, since parameters $\hat{\beta}_3$ and $\hat{\beta}_4$ themselves describe a linear/parabolic trend over time, all flow-adjusted trends in concentration are constrained to follow a linear/parabolic trajectory.

The WRTDS model (Hirsch and others, 2010) also estimates concentration as a log-linear function of discharge, season, and time. The form of the model is

$$\ln(c) = \hat{\beta}_0 + \hat{\beta}_1 t + \hat{\beta}_2 \ln(q) + \hat{\beta}_3 \sin(2\pi t) + \hat{\beta}_4 \cos(2\pi t) + \varepsilon \quad (4)$$

where

\ln	is the natural log function;
c	is the measured concentration, in milligrams per liter;
q	is measured daily-mean discharge, in cubic feet per second;
t	is time, in years;
$\hat{\beta}$	are coefficients;
$\hat{\beta}_0$	is a constant;
$\hat{\beta}_1$	describes the relation between concentration and time;
$\hat{\beta}_2$	describes the relation between concentration and flow;
$\hat{\beta}_3, \hat{\beta}_4$	describe seasonal variation in concentration; and
ε	is the unexplained variation.

The model shown in equation 4 is similar in structure to that used in ESTIMATOR, shown in equation 1; the equations differ in that the WRTDS model lacks the discharge-squared and time-squared terms present in ESTIMATOR. The more significant differences between the two models are the manner in which the coefficients are estimated and the conceptualization and computation of trends. The coefficients in the ESTIMATOR model (eq. 1) are constants, estimated to produce the best fit to the entire sample of measured concentration. In contrast, the coefficients in the WRTDS model (eq. 4)

vary over time. The coefficients are fit by weighted regression, wherein the coefficients for a given day for which an estimate is required are based more heavily on observations collected under conditions that are “close” to those on the day for which an estimate is required. The “closeness” is determined in three dimensions: time, discharge, and season. Concentration values observed within a year or two before or after the date for which an estimate is required influence the regression for that day more than concentration values observed 8 or 9 years before or after the date. Similarly, concentrations observed on days with discharge within a log cycle or two of the discharge on the date in question carry more weight than observations collected under very different discharge conditions, and values collected during about the same time of year influence the regression for a given date more than observations collected in a different season (Hirsch and others, 2010). Even though equation 4 has fewer terms than equation 1, the variable-parameter weighted-regression approach confers greater flexibility in fitting observed water-quality records. Some specific advantages for concentration and flux estimation include the ability to accurately represent (1) concentration-discharge relations that change shape, seasonally or over time, (2) time trends other than parabolic, in particular trends with more than one inflection point, and (3) seasonal patterns that are not strictly sinusoidal, may be different in different discharge ranges, and may change shape over time.

The ability of WRTDS to represent complex relations between concentration and discharge, or complex trends over time, is derived mainly from variability in the model coefficients, which in turn is derived from the weighted-window estimation scheme. As a consequence, trends in the data that are adjusted for flow, analogous to the flow-adjusted trend in concentration computed by ESTIMATOR, cannot be inferred directly from the regression coefficients. Hirsch and others (2010) describe an alternate stochastic framework whereby the set of observed daily flows on any given calendar day over the period of record is taken to define a distribution of possible flow values for that day, with each value equally likely. The mean of the set of concentration values determined by applying equation 4 with time and season fixed on a *particular* day of the record (for example, January 1, 1998) repeated for *all* the daily discharge values observed on that *calendar day* over the simulation period (all values observed on January 1) is interpreted as a point estimate of concentration for that day, independent of the effect of flow. This value is referred to as the *flow-normalized concentration* for January 1, 1998; see Hirsch and De Cicco (2014) for details and qualifying assumptions. Similarly, the *flow-normalized flux* for January 1, 1998, is the mean of the product of the concentrations estimated as above and the flow at which each concentration was estimated; this is taken as a point estimate of flux for January 1, 1998, independent of the effect of flow. Averages of these flow-normalized concentrations (or flow-normalized fluxes) are then computed for each year in the record. *Trends* in flow-normalized values are essentially the difference in flow-normalized values between 2 years,

expressed in either concentration/flux units or as a percentage change relative to the starting value.

A significant benefit of the trend computation used in WRTDS is the ability to estimate independent flow-normalized trends for both concentration and flux. Because they are based on the average of concentration values estimated across the entire range of daily discharges observed on each calendar day, flow-normalized trends in concentration reflect changes in concentration at the most frequently occurring discharge values. In contrast, because they are based on changes in the product of a set of concentration values for any day and the discharge at which they are estimated, flow-normalized trends in flux are often driven by changes in concentration at the highest discharge values, which have a much larger influence on flux than do the days that have discharge values that are near the median discharge or lower (Hirsch and others, 2010).

One practical consequence of the computationally intensive method for estimating trends over time described above for WRTDS is that concentrations must be estimated at many different discharges for every day of the record. One concentration estimate, at the observed daily discharge for any particular day, is required for estimating the concentration and flux associated with that day. For the flow-normalization process described above, concentration estimates are also required for all discharge values that occurred on that calendar day over the entire period of record. Thus, for a 30-year record, and disregarding leap years for illustrative purposes only, $30 \times 365 = 10,950$ estimates are needed for estimating the actual flux/concentration on each day of the record, and $29 \times 30 \times 365 = 317,550$ additional estimated values are needed for flow normalization. Rather than applying equation 4 at all of the specific points in the time-discharge domain where an estimated concentration is required, values are estimated at all points on a two-dimensional grid spanning the period of record (the “x” dimension) and the full range of discharge values observed over the period of record (the “y” dimension). By current convention, the points along the “x” dimension are computed by dividing each year of the simulation into 16 equal increments; similarly, points along the “y” dimension are computed by dividing the range of observed discharges, in logarithmic space, into 14 equal increments. Thus, in the example above, equation 2 is applied at only $30 \times 16 \times 14 = 6,720$ points. The result is a smoothly varying surface of estimated concentration (the “z” dimension), from which the specific concentrations required for flux estimation and flow normalization are interpolated. In addition to saving considerable computational time, the estimated concentration surface, viewed as a contour map, can itself be visually interpreted as one means of assessing, for example, the evolution of the concentration-discharge relation over time. Utility routines for directly viewing this surface are included in the WRTDS distribution (Hirsch and De Cicco, 2014).

Another consequence of the weighted-regression, variable-coefficient approach used in WRTDS is the difficulty, relative to constant-coefficient models such as ESTIMATOR, of specifying uncertainty bounds on flux and trend estimates.

The model was originally presented as a tool for exploratory analysis (Hirsch and others, 2010) and continues to be useful as such (Murphy and others, 2014; Corsi and others, 2015). As of this writing, however, concentrations, fluxes, and trends are reported by WRTDS without uncertainty bounds and, as such, are not suitable for statistical hypothesis testing. Provision of uncertainty bounds for WRTDS results is currently an active topic of research.

Previously Published Comparisons of ESTIMATOR and WRTDS

Moyer and others (2012) published the first report specifically focused on comparing ESTIMATOR and WRTDS. The main focus of the study was to compare the accuracy and bias of the two models' estimates of annual flux for five constituents: total nitrogen, nitrate, total phosphorus, orthophosphate, and suspended sediment. The study considered data from the nine Chesapeake Bay RIM stations (fig. 1; table 1) for water years 1985 through 2010. These stations have some of the longest records and densest sampling frequencies in the CBNTN; moreover, water-quality sampling at these stations has consistently been stratified to emphasize sampling infrequent high flows (Langland and others, 2006). As a result, water-quality records collected at these stations are among the most suitable of all records in the CBNTN for regression-based flux and trend estimation.

Flux comparisons by Moyer and others (2012) were based primarily on model explanatory power, or the degree to which the model fits the data, and flux bias, the tendency of the model to over- or underpredict flux overall. Specifically, Moyer and others (2012) quantified model explanatory power in terms of the root-mean-square error (RMSE) of the model residuals and flux bias as the ratio of the sum of the estimated fluxes on sampled days to the sum of observed fluxes on the same days. For all five constituents at all nine stations, WRTDS had a lower residual RMSE than ESTIMATOR, with at least three of the nine stations showing improvement of 20 percent or more for all five constituents. Broadly speaking, the degree of reduction in RMSE was greatest for the two particulate constituents: suspended sediment and total phosphorus (frequently dominated by the particulate fraction). WRTDS models also proved to be less biased than ESTIMATOR for 37 of the 45 station-constituent combinations; for 16 combinations WRTDS results had a flux bias ratio closer to 1.0 (indicating no bias) than ESTIMATOR by a factor of 0.10 or greater. The degree of improvement in flux bias, as well as the number of stations where improvement was observed, was greatest for suspended sediment and least for total nitrogen. Moyer and others (2012) explained the improvement gained with WRTDS primarily in terms of that model's flexibility, relative to ESTIMATOR, in representing concentration-discharge relations that were not purely linear or parabolic.

Trend comparisons at the nine RIM stations (Moyer and others, 2012) were focused in particular on comparing trends in flow-adjusted concentration generated by ESTIMATOR to trends in flow-normalized yield, or flux divided by basin area, generated by WRTDS. The comparison was conducted primarily to introduce the scientific and regulatory communities, accustomed to the well-established notion of a trend in concentration that is adjusted for discharge, to the relatively novel concept of a similarly adjusted trend in flux. Strict statistical hypothesis testing was hindered by the absence of confidence bounds on WRTDS trend estimates. For trends estimated over the entire period of record, however, Moyer and others (2012) found that the concentration and yield trends differed in sign for at least one of nine stations for all constituents except total nitrogen. The long-term record for total phosphorus at the James River at Cartersville, VA (JAMC, USGS station 02035000), was used as a case study of this phenomenon. At that station, an overall downward trend in concentration at the most frequently occurring discharges led to a downward trend in flow-adjusted concentration, while an upward trend in concentration at the highest discharges led to an upward trend in flow-normalized flux. Although the study of Moyer and others (2012) explained the mechanisms by which flow-adjusted concentration and flow-normalized flux trends could differ in sign, it did not fully address several questions related to the fundamental differences in trend estimation between the models: First, why is ESTIMATOR limited to producing a concentration trend but not a flux trend? Second, given that the means by which concentration trends themselves are estimated differ substantially between ESTIMATOR and WRTDS, how closely should the two estimates agree? If the estimates differ substantially, which method is more credible? These questions are addressed in detail in the current report, in the subsections "Trends in Concentration" and "Trends in Flux" under the section "Comparison of ESTIMATOR and WRTDS Models on Chesapeake Bay Nontidal Water-Quality Monitoring Network Data Through Water Year 2012."

Hirsch (2014) conducted a detailed study of the susceptibility of regression-based flux estimation methods to flux bias, using six densely sampled 10-year water-quality records from the upper Midwest. Estimates of 10-year average flux, based on subsamples from a nearly daily baseline record of concentration observations, were compared to the "true" 10-year flux, determined using the dense baseline record. The subsamples were drawn to reflect differing overall sampling frequencies, as well as differing levels of preference for high-flow samples. Hirsch (2014) found that the bias of WRTDS flux estimates was almost always lower than a model equivalent to ESTIMATOR and that differences were small for the exceptions. This pattern was evident at all sample sizes considered, although subsampling variability in flux bias increased with decreasing subsample size. On this basis, Hirsch (2014) suggests that in broad application where the number of observations exceeds 120, the likelihood of severely biased results is smallest with WRTDS.

Discussing these results, Hirsch (2014) outlines a set of four features of water-quality records, any of which can lead to bias problems in flux estimation models with fixed coefficients (such as ESTIMATOR). To paraphrase, these features are (1) concentration-discharge relations that, though they may remain consistent over time and season, have overall shapes not well represented as linear or quadratic functions in log-log space, (2) severely heteroscedastic (nonconstant-variance) residuals, (3) concentration-discharge relations that change shape seasonally, and (4) concentration-discharge relations that change shape over time. Hirsch (2014) stresses that all regression-based approaches to estimating daily concentrations, based on a sparse set of water-quality observations, commonly suffer to some degree from one or more of the issues listed above. Careful examination of regression diagnostics can often reveal the problematic cases; Hirsch (2014) illustrates the use of a panel of eight regression diagnostic plots aimed in particular at identifying problems arising from the four features listed above. A function for producing this panel of plots is included in the WRTDS distribution (R package “EGRET”; Hirsch and De Cicco, 2014).

To draw a link between a regression model’s tendency to over- or underestimate daily fluxes on sampled days and the potential for over- or underestimating the true aggregated 10-year flux, determined directly through dense sampling in Hirsch’s (2014) study but generally unknown, the author proposed a *flux bias statistic*, B , defined as

$$B = (P - O) / P \quad (5)$$

where

$$O = \sum_{i=1}^n c_i \cdot Q_i \quad (6)$$

and

$$P = \sum_{i=1}^n \hat{c}_i \cdot Q_i \quad (7)$$

where O is the sum of observed fluxes on sampled days, P is the sum of estimated fluxes on sampled days, c_i are the observed concentrations, Q_i are the observed discharges, and \hat{c}_i are the estimated concentrations. Hirsch (2014) calculated this statistic for every model used in the study. Against that statistic, the author plotted that model’s true relative error in 10-year average flux (hereafter, “true relative error”) computed as the difference in the estimated 10-year average flux and the true flux on the basis of dense sampling and expressed as a percentage of the true 10-year average flux. Flux bias statistics less than about -0.1 were associated with true relative errors between -20 percent and 0 ; however, there was little relation between the flux bias statistic and the true relative error over that range. Flux bias statistics between about -0.1 and $+0.1$ were indicative of true relative errors between -10 and $+10$ percent; again, however, the statistic was a poor predictor of the true relative error. As the flux bias statistic increased from about $+0.1$ to $+0.4$, true relative errors

increased over roughly the same range expressed in percentage points (that is, $+10$ to $+50$ percent), although there was considerable scatter in the relation. Flux bias statistics greater than about $+0.4$ were associated with rapidly increasing true relative errors: a statistic of $+0.5$ corresponded roughly to a true relative error of $+100$ percent and statistics around $+0.7$, the highest encountered in the study, corresponded to true relative errors on the order of 200 percent (Hirsch, 2014).

Comparison of ESTIMATOR and WRTDS Models on Chesapeake Bay Nontidal Water-Quality Monitoring Network Data Through Water Year 2012

Moyer and others (2012) and Hirsch (2014) compared the performance of WRTDS and ESTIMATOR using datasets selected specifically for their length, density of sampling, and consistency in sampling protocol. As a first step in comparing the relative performance of the two models for use in an operational water-quality network such as the CBNTN, both models are now applied to selected groups of watersheds for which flux and (or) trend results were reported for total nitrogen, total phosphorus, and sediment through water year 2012.

Description of Scenarios Considered

For purposes of understanding which datasets were used for some of the analyses discussed and recommendations made in this section, it is helpful to be aware of the distinction between how flux and trend results through water year 2012 were reported on the CBNTN Web site and how they were computed internally. The suite of numeric products currently provided online (<http://cbrim.er.usgs.gov/index.html>) includes annual fluxes, 5-year average annual yields (flux per unit area) for the most recent 5-year period, long-term (period-of-record) trends for stations having greater than 20-year records, and 10-year trends for stations having at least 10-year records. Only fluxes and yields (no trends) are reported for stations having less than 10-year records, and no results of any kind are reported for stations having less than 5-year records. As discussed in the “Introduction” section, WRTDS was introduced in 2012 for the computation of fluxes and yields, but ESTIMATOR was retained for estimation of trends. Both legacy practices and the novelty of introducing a second model dictated that model runs to produce the 2012 results be divided internally into five main groups, or “scenarios”:

1. *Stations which had water-quality records exceeding 20 years in length*, referred to hereafter as “*LONG TERM*”; these include the nine RIM stations and, for most constituents, 21 other stations. Models for these stations were estimated using water-quality and discharge data spanning each station’s period of record, typically between 25 and 30 years (fig. 2; table 1).

Models were estimated using both WRTDS, for fluxes and yields, and ESTIMATOR, for trends over the period of record.

2. *Stations which had water-quality records of at least 10 years' length*, hereafter, "*10_YEAR*"; these include all the stations in no. 1 above, as well as a few other stations not satisfying the 20-year criterion for *LONG_TERM* records. In this scenario, records for any station having more than 10 years' data were truncated to the 10-water-year period 2003 through 2012. All stations were then modeled using WRTDS, but fluxes and yields from these runs were only reported for those stations not satisfying the *LONG_TERM* criterion; that is, fluxes and yields were reported from *LONG_TERM* runs wherever they were available. All stations were also modeled using ESTIMATOR (for reporting 10-year trends).
3. *Fifteen Virginia stations having records of at least 10 years, but with only monthly data (no storm sampling)*, hereafter, "*10_YEAR_SECONDARY*" (table 1). This is the only group of stations not having both monthly and targeted storm sampling. Records for these stations were truncated to the 10-water-year period 2003 through 2012. No fluxes or yields, only 10-year trends modeled using ESTIMATOR, were reported.
4. *Stations having between 5 and 10 years' data and having a total of 120 or more observations*, hereafter, "*LESS_THAN_10_WRTDS*." WRTDS was used to calculate fluxes and yields for these stations; trends were not reported due to the stations not having the minimum 10 years' data. The distinction between this and the following group was made on the basis of the lack of published testing of WRTDS, at the time results were posted, on datasets with fewer than 120 observations.
5. *Stations having between 5 and 10 years' data and having fewer than 120 observations*, hereafter, "*LESS_THAN_10_ESTIMATOR*." Because the total number of samples for these stations fell below the lower limit of any published studies at the time water year 2012 results were posted, both fluxes and yields for these stations were modeled using ESTIMATOR; these are the only stations for which ESTIMATOR was used as the basis for reporting fluxes and yields. Because record lengths differed between constituents, some stations fell under scenario *LESS_THAN_10_WRTDS* for some constituents and *LESS_THAN_10_ESTIMATOR* for others.

The short names listed above will be used in discussing some of the analyses and recommendations that follow in this report.

Results are reported on the CBNTN Web site for a total of five constituents: total nitrogen, dissolved inorganic nitrogen (nitrate plus nitrite), total phosphorus, orthophosphate, and sediment. Because some member jurisdictions have historically monitored total suspended solids (TSS) rather than suspended-sediment concentration (SSC), the particular

constituent reported as "sediment" varies from station to station (table 1; see Glysson and others (2001) for a discussion of the comparability of TSS and SSC data). For brevity, only results for total nitrogen, total phosphorus, and sediment are compared and discussed in this section of the current report. WRTDS-calculated annual results through water year 2012 for all stations and all constituents are provided in appendix 3.

All results reported, and comparisons made, throughout this report were obtained using the version of WRTDS implemented in the EGRET package, version 1.2.4, with supporting routines from the dataRetrieval package, version 1.2.2 (<https://github.com/USGS-R/EGRET/wiki>), running under R x64 version 3.0.1 (R Core Team, 2013). WRTDS was used with default parameters for half-window width (10 years for time, 2 log cycles for discharge, and 6 months for season) and minimum numbers of observations (100 overall, 50 uncensored). Note that since the time that these results were run, the version of WRTDS implemented in EGRET (version 2.0 and later, available through the Comprehensive R Archive Network; <http://cran.us.r-project.org/>) has been modified to improve the stability of estimates in the first and last few years of the record (the "edgeAdjust" feature; Hirsch and De Cicco, 2014). One component of this modification is the adoption of a default time half-window width of 7 years.

Measures of Model Performance

This section compares the performance of ESTIMATOR and WRTDS on data from an operational water-quality network using two common means of characterizing the performance of regression-based flux-estimation models: explanatory power and flux bias. The comparisons that follow were performed on the three CBNTN model scenarios for which WRTDS flux results were originally reported for water year 2012: *LONG_TERM*, *10_YEAR*, and *LESS_THAN_10_WRTDS*. Because ESTIMATOR flow-adjusted trend results were originally reported for the first two of these scenarios, ESTIMATOR model results and performance statistics were already available. The ESTIMATOR model was run on the *LESS_THAN_10_WRTDS* scenario expressly for purposes of conducting the comparisons reported in this section. Overall, WRTDS model performance statistics were consistently found to be at least equal to, and in some cases distinctly better than, corresponding statistics for ESTIMATOR models fit to the same data.

Explanatory Power

Moyer and others (2012) compared the relative effectiveness of WRTDS and ESTIMATOR in explaining variability of a set of water-quality observations about the set's mean concentration using the root-mean-square error (RMSE). This is an appropriate technique for comparing two different modeling approaches on data having similar ranges of values. However, because different constituents have different concentration ranges and thus different degrees of variability,

the metric is unsuitable for comparing model performance over a suite of different constituents. In the following comparison, explanatory power is rated in terms of the proportion of the sum of a set of observations' squared deviations from its grand mean that is explained by the model. When the data contain no uncensored values, this is equivalent to the R^2 statistic commonly used in regression modeling. When the data contain censored observations, a strict partitioning of variability in this manner is problematic. For the comparisons that follow, censored observations are replaced by their "average" concentration, defined as follows: if the observation is interval-censored, the "average" concentration is taken as the mid-point of the interval; for the results considered here, only the nine RIM stations have interval-censored observations. If the observation is left-censored, the "average" concentration is taken to be one-half of the implied reporting limit (that is, if the concentration is reported as "< 0.2 milligrams per liter [mg/L]," then a value of 0.1 mg/L is used in the computation). This approximation, applied uniformly to all constituents, is suitable for the coarse comparisons considered in this section.

For all three scenarios considered, explanatory power for both models was lowest for total nitrogen and highest for sediment (fig. 3). For example for the *LESS_THAN_10_WRTDS* scenario, median values for both models pooled were 48.2 percent for total nitrogen, 62.7 percent for total phosphorus, and 67.9 percent for sediment (fig. 3C). This likely reflects the varying degrees to which each constituents' concentration is influenced by discharge. Both nitrogen and phosphorus are cycled biogeochemically in both terrestrial and aquatic environments. Although discharge plays a role in both the concentration and speciation of these constituents, the total concentration of nitrogen and phosphorus, both at any given station and between stations, is also influenced by this cycling. Sediment, on the other hand, is mobilized both on land and within the stream channel almost exclusively by energy associated with falling or flowing water. Total phosphorus concentrations throughout the CBNTN are typically dominated by the particulate fraction, while the speciation of nitrogen is more variable; this is the likely explanation for the model explanatory power for total phosphorus lying closer to that found for sediment. The explanatory power of any given model also tends to be lowest for the *LONG_TERM* scenario and highest for the *LESS_THAN_10_WRTDS* scenario (for example, for sediment, the pooled-model median explanatory power for the three scenarios, in order of decreasing record length, was 59.4, 66.2, and 67.9 percent, respectively). This is probably due to the fact that shorter records have essentially no time trend component and generally represent a narrower range of environmental conditions and thus fewer potential sources of unexplained variability.

With regard to the relative performance of WRTDS and ESTIMATOR, it is perhaps most noteworthy that the overall differences in explanatory power are not large (fig. 3); which is evidence of the fundamental challenges and limitations of all regression methods based on discharge (for example,

Hirsch, 2014). For the shortest records (*LESS_THAN_10_WRTDS*; fig. 3C), the performance of the two models is very similar for all three constituents; a small degree of overall improvement in the median explanatory power is evident for *10_YEAR* datasets (fig. 3B). A distinct tendency for WRTDS to account for more variability than ESTIMATOR is only evident in the *LONG_TERM* records (fig. 3A), where median model explanatory power improves by roughly 6 percent for total nitrogen (33.9 to 39.9 percent), 8 percent for total phosphorus (38.6 to 46.9 percent) and 5 percent for sediment (56.5 to 61.2 percent). Recalling that WRTDS operates with a minimum 10-year time half-window and weights observations within that window, the small improvement in WRTDS' performance, relative to ESTIMATOR, between the *LESS_THAN_10_WRTDS* and *10_YEAR* scenarios probably reflects a transition from a situation where the benefit of time-weighting provided by WRTDS represents only a small improvement over the unweighted approach used by ESTIMATOR to one where weighting in time begins to confer a more noticeable advantage (that is, for a 5-year record, the influence of weighting is rather small relative to the influence in a record of 10 years or longer). The greater improvement on long-term datasets reflects the additional benefit of "sliding" the 10-year half-window through time, allowing for more flexible trend representation and the ability to account for concentration-discharge relations that change shape over time.

Flux Bias

Flux bias statistics, as defined by Hirsch (2014) and computed using equation 5, were also compared using the *LONG_TERM*, *10_YEAR*, and *LESS_THAN_10_WRTDS* scenarios, with censored values treated as described above. For all three scenarios, flux bias statistics were lowest for total nitrogen and highest for suspended sediment (fig. 4). For example, pooling the statistic for both models for the *LONG_TERM* scenario, the median flux bias values were -0.005 for total nitrogen, +0.033 for total phosphorus, and +0.149 for sediment (fig. 4A). This pattern likely results from basic differences in the concentration-discharge relations for these constituents. Log-transformed sediment concentrations typically increase, albeit along different trajectories from station to station, with increasing log discharge; the same is generally true for total phosphorus. However, this relation frequently breaks down or even reverses itself at the extreme upper end of the range of sampled discharges; this could be due to flood-plain deposition, concentration-discharge hysteresis effects, limits to the supply of material subject to suspension, or even methodological issues associated with sampling overbank flows. The failure of either ESTIMATOR or WRTDS to accurately represent this reversal results in lack-of-fit, specifically overprediction, at the highest discharges. Because flux is the product of concentration and discharge, and because the flux bias statistic is calculated in arithmetic space, even modest overprediction of one or two concentration observations at extreme discharges can result in a model with an overall positive bias. Consistent with this mechanism and with

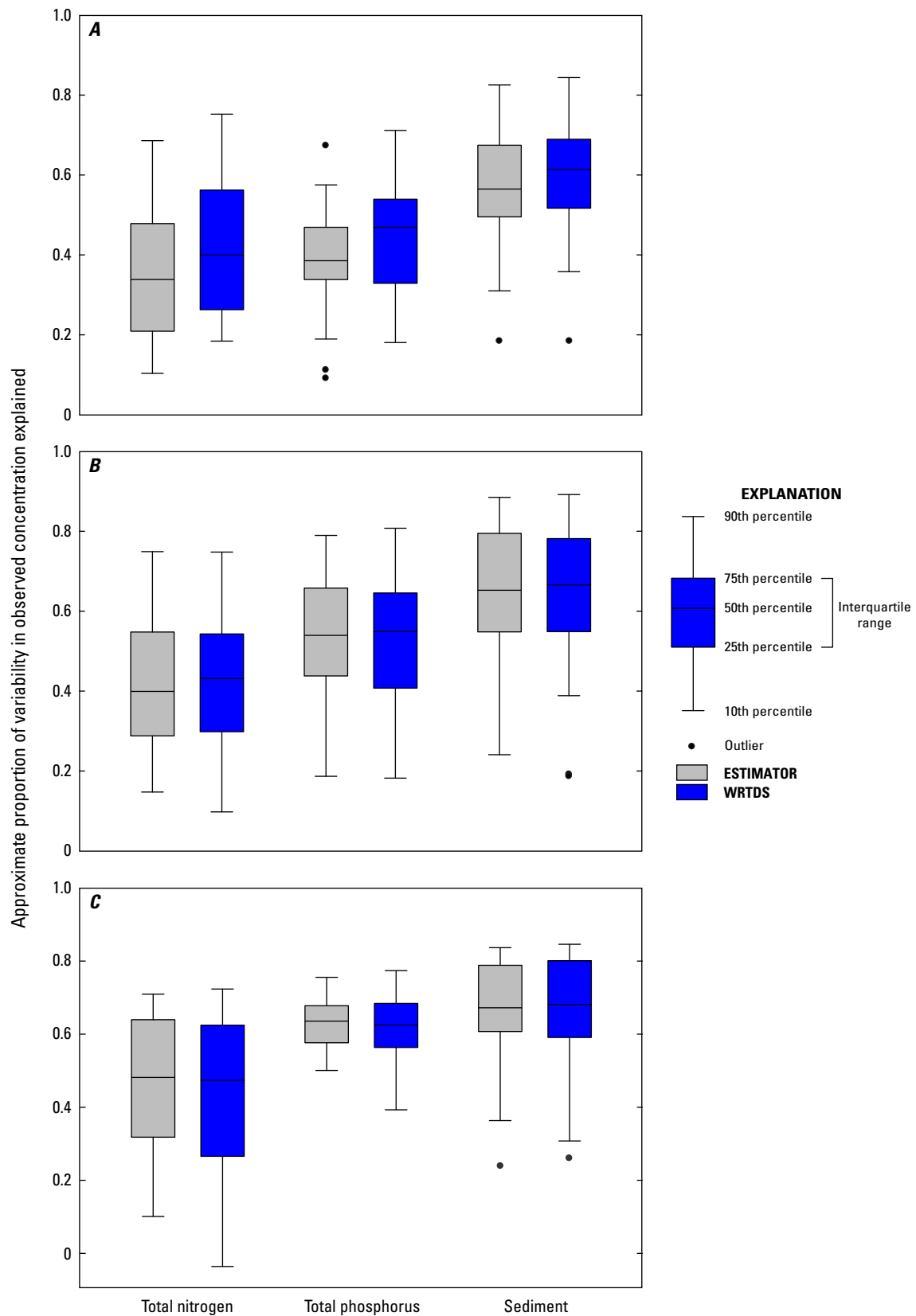


Figure 3. Boxplots comparing model explanatory power between ESTIMATOR and WRTDS (Weighted Regressions on Time, Discharge, and Season) models for the three Chesapeake Bay Nontidal Water-Quality Monitoring Network (CBNTN) scenarios for which WRTDS flux results were reported in water year 2012. A, *LONG_TERM* scenario, B, *10_YEAR* scenario, and C, *LESS_THAN_10_WRTDS* scenario.

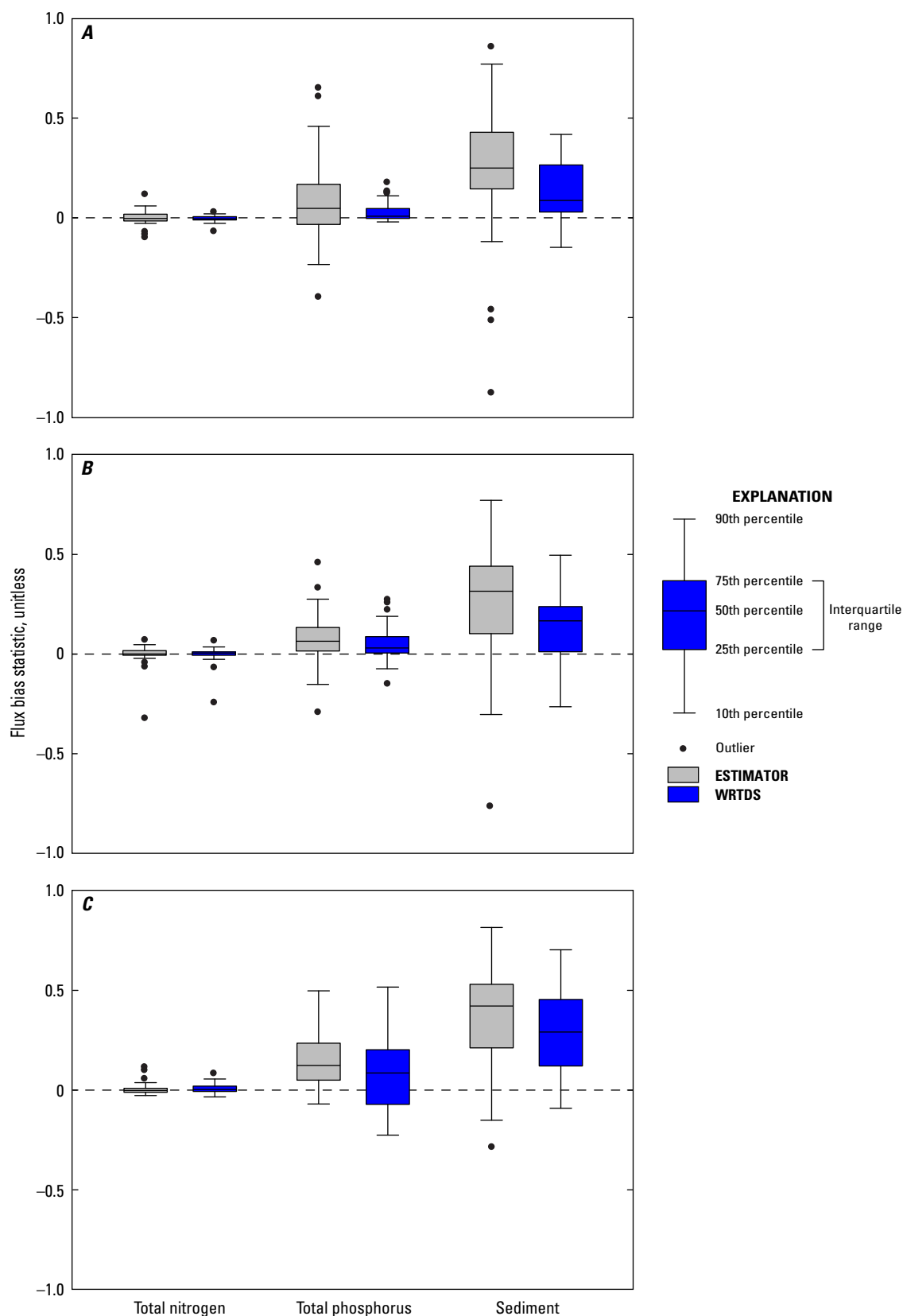


Figure 4. Boxplots comparing flux bias statistics between ESTIMATOR and WRTDS (Weighted Regressions on Time, Discharge, and Season) models for the three Chesapeake Bay Nontidal Water-Quality Monitoring Network (CBNTN) scenarios for which WRTDS flux results were reported in water year 2012. *A*, *LONG_TERM* scenario, *B*, *10_YEAR* scenario, and *C*, *LESS_THAN_10_WRTDS* scenario.

the findings of Moyer and others (2012) and Hirsch (2014), the more flexible model formulation of WRTDS confers the biggest advantage with respect to bias in the CBNTN datasets with sediment: for the *LONG_TERM* scenario, median flux bias statistics differed very little between the two models for total nitrogen and total phosphorus, but for sediment the statistics improved (decreased) from +0.249 for ESTIMATOR to +0.087 for WRTDS (fig. 4A). Similar between-constituent patterns of improvement were evident for the *10_YEAR* and *LESS_THAN_10_WRTDS* scenarios. In contrast to the case for model explanatory power, either model's performance for any constituent tended to degrade, that is, bias statistics increased, with shorter record lengths: taking sediment as an example, pooled median flux bias statistics were +0.149 for the *LONG_TERM* scenario, +0.203 for the *10_YEAR* scenario, and +0.340 for the *LESS_THAN_10_WRTDS* scenario (fig. 4). This results from the fact that the statistic represents an average over all observations, and poorly predicted individual observations have greater influence when the overall count is smaller.

Estimated Yields

In this section, model estimates of 5-year average annual yield (flux per unit basin area) for total nitrogen, total phosphorus, and sediment are compared. This comparison was first performed, using scatter plots, for the same scenarios used to compare model performance statistics in the previous section: *LONG_TERM*, *10_YEAR*, and *LESS_THAN_10_WRTDS*. However, the 5-year yield comparison for the *LONG_TERM* scenarios was broadly representative of results obtained from the other two shorter-length scenarios; thus, only results from the *LONG_TERM* scenario are discussed below. Broad patterns of differences in the two models' estimates of yield across the CBNTN are found to be consistent with previously documented differences in the models' ability to handle constituent-specific features of the observed concentration-discharge relations.

A comparison between ESTIMATOR and WRTDS estimates of 5-year (2008–12) annual average yields for total nitrogen, total phosphorus, and suspended sediment for stations run under the *LONG_TERM* scenario is shown in figure 5. Yields generally plotted around the 1:1 line for the

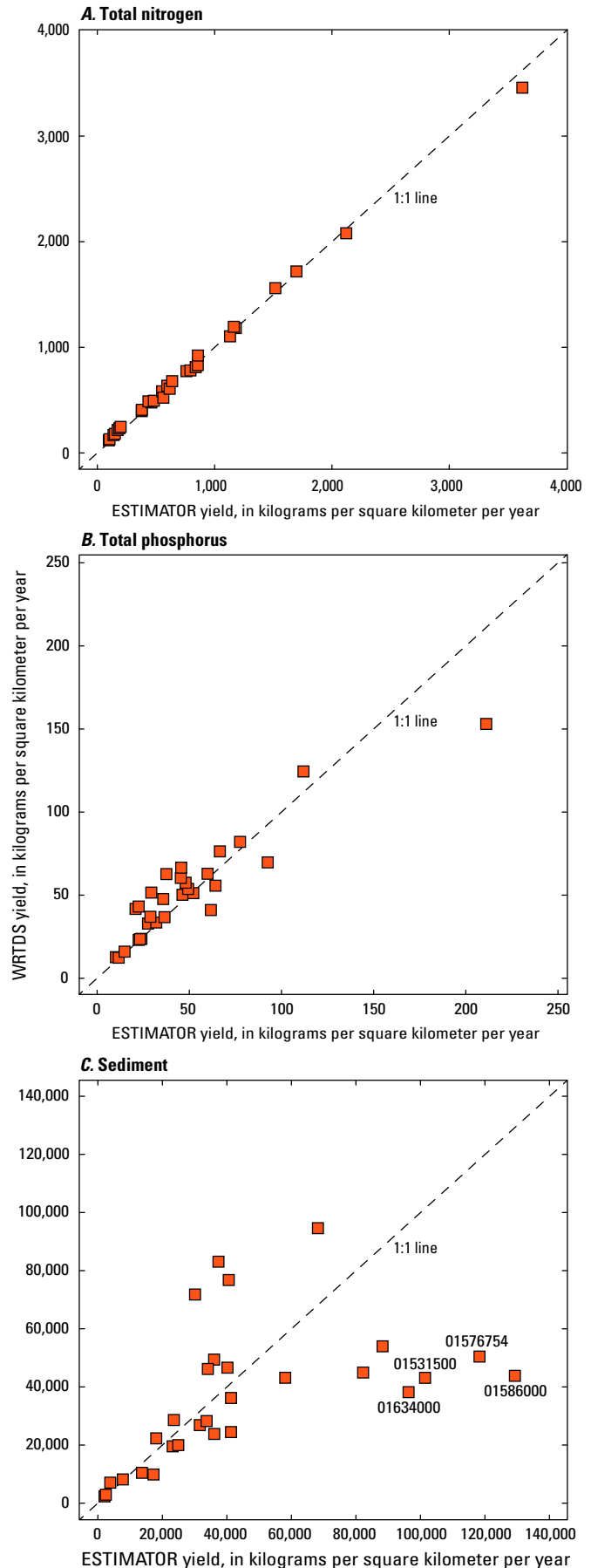


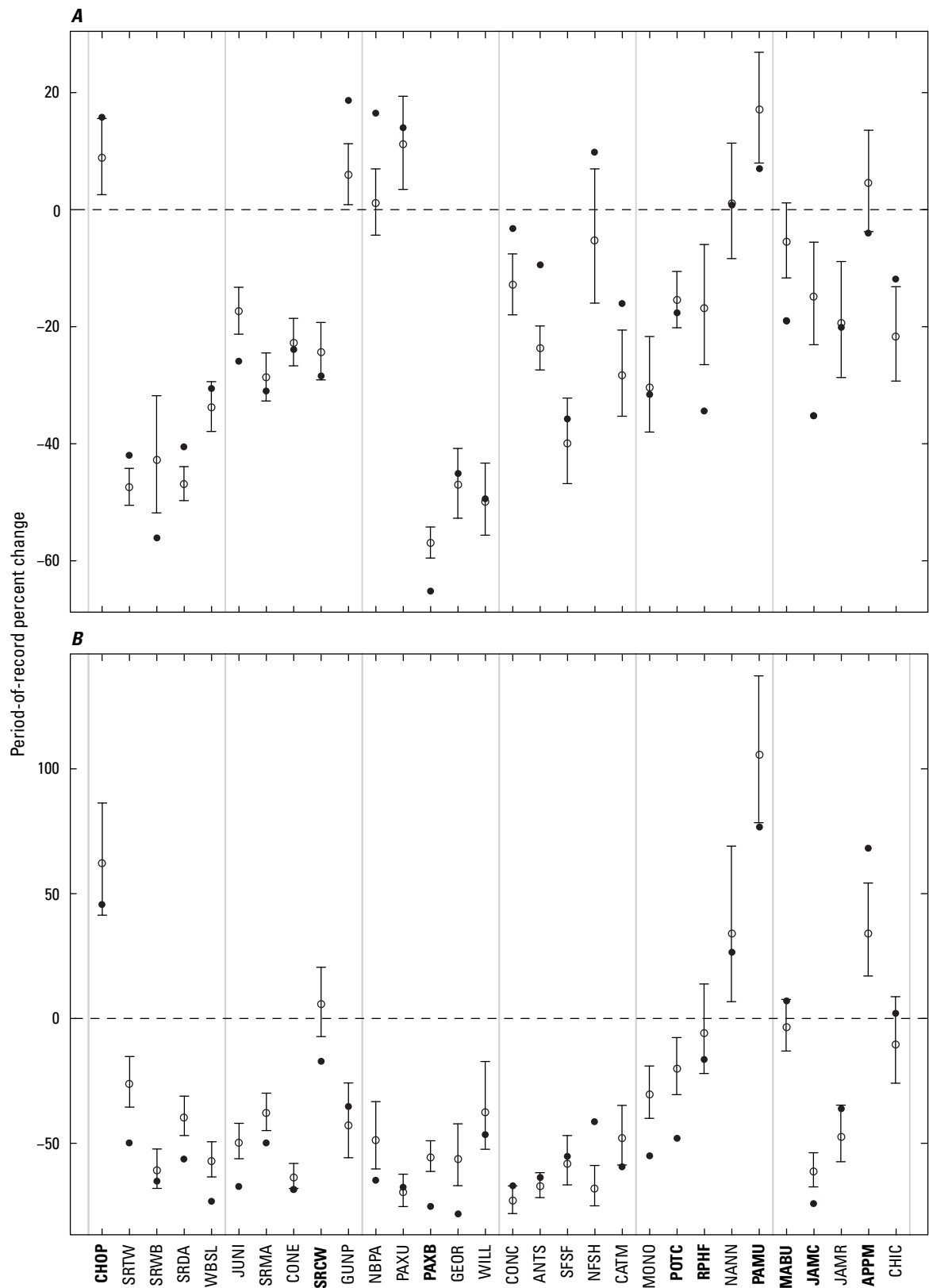
Figure 5. Scatter plots comparing estimated 5-year (water year 2008–12) average yields of nutrients and sediment computed using ESTIMATOR and WRTDS (Weighted Regressions on Time, Discharge, and Season) for the Chesapeake Bay Nontidal Water-Quality Monitoring Network (CBNTN) 2012 *LONG_TERM* scenario. A, Total nitrogen, B, total phosphorus, and C, sediment. Numbered stations are identified in table 1.

three constituents, with the least scatter for total nitrogen and the most scatter for suspended sediment. The between-model agreement in results for total nitrogen stems from the fact that of the three constituents, total nitrogen has the weakest relation with discharge; although WRTDS provides a small advantage in explanatory power over ESTIMATOR for total nitrogen (fig. 3A), there is no clear difference in either models' tendency to under- or overpredict observations (fig. 4A). In contrast, there is considerable scatter about the 1:1 line for suspended sediment yield (fig. 5C), with an overall broader range of yield estimates for ESTIMATOR and a tendency for ESTIMATOR to predict higher values than WRTDS for the highest yields. The larger overall degree of scatter is consistent with the tendency for ESTIMATOR to have a broader distribution of flux bias statistics than WRTDS, including both extreme positive and extreme negative values (fig. 4A). The four points closest to the right side of figure 5C depict sediment yields at the Conestoga River at Conestoga, PA (CONE, USGS station 01576754), the North Branch Patapsco River at Cedarhurst, MD (NBPA, USGS station 01586000), the Susquehanna River at Towanda, PA (SRTW, USGS station 01531500), and the North Fork Shenandoah River near Strasburg, VA (NFSH, USGS station 01634000); concentration-discharge plots for these stations have relations that are parabolic-upward throughout most of the discharge range, but most observations at the highest discharges fall below the fitted parabola (data not shown). These stations are in the "Category 2" class defined by Moyer and others (2012), a pathology in which the upward-parabolic approximation of the concentration-discharge relation produced by ESTIMATOR overestimates, in some cases by a wide margin, sediment concentrations at high discharges relative to the more flexible relation fitted by WRTDS. This situation is reflected in the relative flux bias statistics for the four stations listed above, which range from 0.29 to 0.86 for ESTIMATOR and 0.05 to 0.31 for WRTDS.

Trends in Concentration

The previous sections established that, when both WRTDS and ESTIMATOR were applied to a broad range of datasets across the CBNTN, WRTDS performance statistics equaled or exceeded those of ESTIMATOR and that between-model differences in estimated yield were consistent with previous findings, indicating that the flexibility of a weighted-regression approach conferred an advantage in modeling specific, commonly occurring peculiarities in concentration-discharge relations. This section considers trends over time, focusing first on summarizing differences between the trend in flow-adjusted concentration reported by ESTIMATOR and the trend in flow-normalized concentration reported by WRTDS across the CBNTN. WRTDS point estimates of trend in flow-normalized concentration of total nitrogen for the *LONG_TERM* CBNTN scenario, expressed as a percentage change between the starting and ending years of the record for each station, are plotted with corresponding ESTIMATOR estimates and confidence intervals of flow-adjusted trends in concentration in figure 6A. In 18 of the 30 trends illustrated, the WRTDS flow-normalized trend point estimate lies outside the confidence interval for the ESTIMATOR flow-adjusted trend estimate. Nevertheless, the WRTDS estimates lie near enough to the ESTIMATOR estimates that the overall pattern of both magnitude and direction of the two trend estimates across all stations is, at least to a degree, preserved. A similar pattern of broad agreement exists for total phosphorus (fig. 6B) and sediment (fig. 6C; trend results for phosphorus and sediment at some stations are not shown, either because the given constituent was not modeled at that station or because the ESTIMATOR model was disqualified due to severe lack-of-fit during diagnostic checks). For the same constituents in the CBNTN *10_YEAR* scenario, all WRTDS point estimates of trend in flow-normalized concentration lie within the confidence intervals of the corresponding ESTIMATOR trend (fig. 7). Taken together, figures 6 and 7 indicate that the two trend estimation methods, although defined and computed quite differently, convey similar information. Making an informed decision as to which model represents the data most faithfully requires a better understanding of the mechanisms underlying these similarities. These mechanisms are explored by means of a specific case study in the following section.

Figure 6. (facing page) Plots comparing the flow-adjusted trend in concentration computed using ESTIMATOR with the flow-normalized trend in concentration computed using WRTDS (Weighted Regressions on Time, Discharge, and Season) for the Chesapeake Bay Nontidal Water-Quality Monitoring Network (CBNTN) 2012 *LONG_TERM* scenario. A, Total nitrogen, B, total phosphorus, and C, sediment. Station abbreviations are identified in table 1.



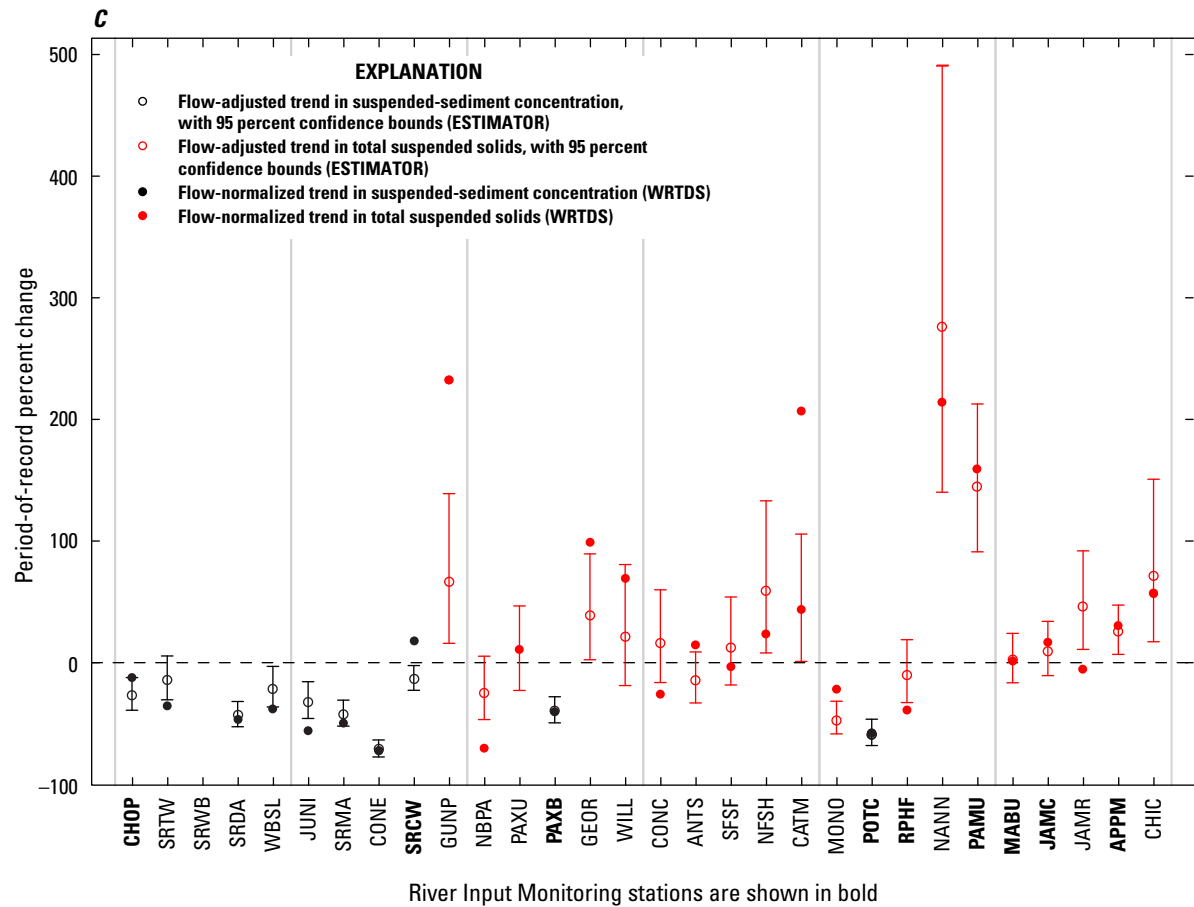
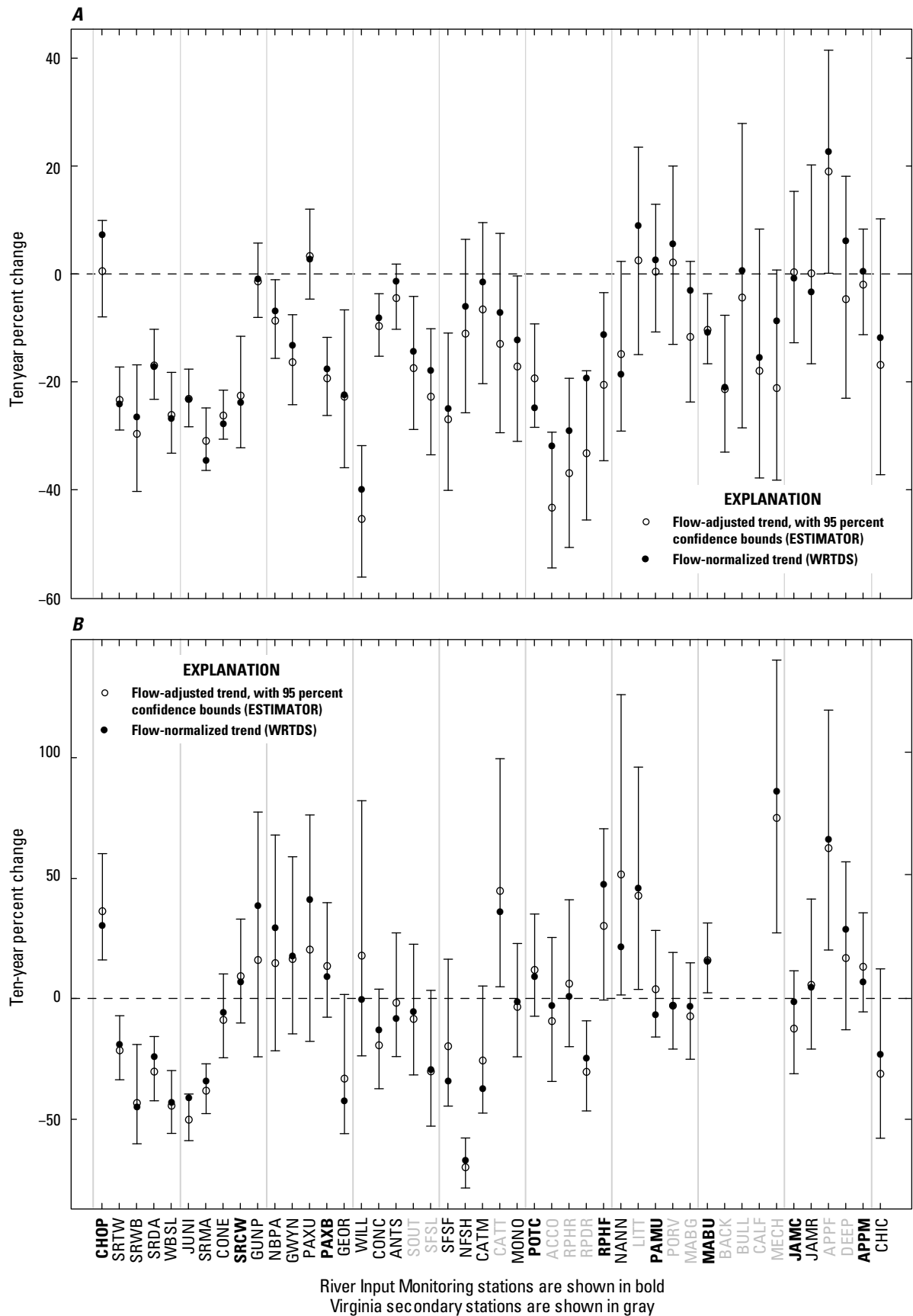


Figure 6. Plots comparing the flow-adjusted trend in concentration computed using ESTIMATOR with the flow-normalized trend in concentration computed using WRTDS (Weighted Regressions on Time, Discharge, and Season) for the Chesapeake Bay Nontidal Water-Quality Monitoring Network (CBNTN) 2012 *LONG_TERM* scenario. *A*, Total nitrogen, *B*, total phosphorus, and *C*, sediment. Station abbreviations are identified in table 1.—Continued

Figure 7. (facing page) Plots comparing the flow-adjusted trend in concentration computed using ESTIMATOR with the flow-normalized trend in concentration computed using WRTDS (Weighted Regressions on Time, Discharge, and Season) for the Chesapeake Bay Nontidal Water-Quality Monitoring Network (CBNTN) 2012 *10_YEAR* scenario. *A*, Total nitrogen, *B*, total phosphorus, and *C*, sediment. Station abbreviations are identified in table 1.



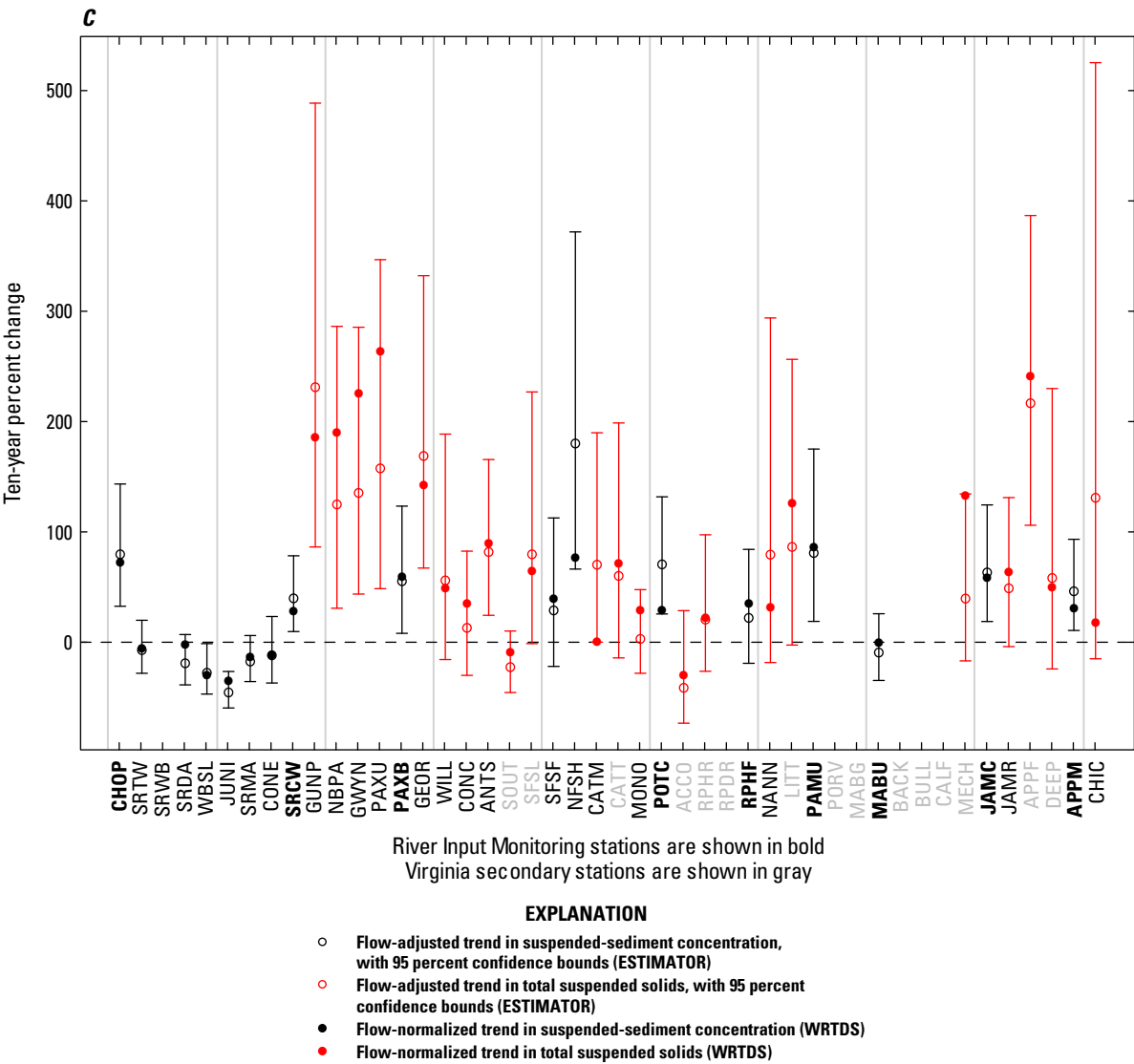


Figure 7. Plots comparing the flow-adjusted trend in concentration computed using ESTIMATOR with the flow-normalized trend in concentration computed using WRTDS (Weighted Regressions on Time, Discharge, and Season) for the Chesapeake Bay Nontidal Water-Quality Monitoring Network (CBNTN) 2012 *10_YEAR* scenario. *A*, Total nitrogen, *B*, total phosphorus, and *C*, sediment. Station abbreviations are identified in table 1.—Continued

Reconciling the Concepts of “Flow-Adjusted Concentration” and “Flow-Normalized Concentration”

Langland and others (2006) describe the flow-adjusted trend in concentration as a technique developed to “...compensate for the influence of flow variability, to better understand changes in concentrations that may be the result of human activities” (p. 17). As described previously in the overview of the ESTIMATOR model, the flow-adjusted trend is computed directly from the coefficients $\hat{\beta}_3$ and $\hat{\beta}_4$ in the model defined in equation 1 and is by convention expressed as a percentage change relative to the concentration at the beginning of the record. Hirsch and others (2010) describe

flow-normalization as a means of providing the user with “[concentration and flux] histories that eliminate the influence of year-to-year variations in streamflow” (p. 857). As summarized in the preceding overview of the WRTDS model, computation of flow-normalized values hinges on a very different stochastic perspective. One potential means of gaining insight as to how these two different approaches yield comparable results for concentration trends (figs. 6 and 7) is to adopt the WRTDS empirical-stochastic approach to flow-normalization but, in doing so, estimate the concentration surface from which values used for flow-normalization are interpolated using ESTIMATOR (that is, equation 1, with coefficients estimated using un-weighted regression) instead

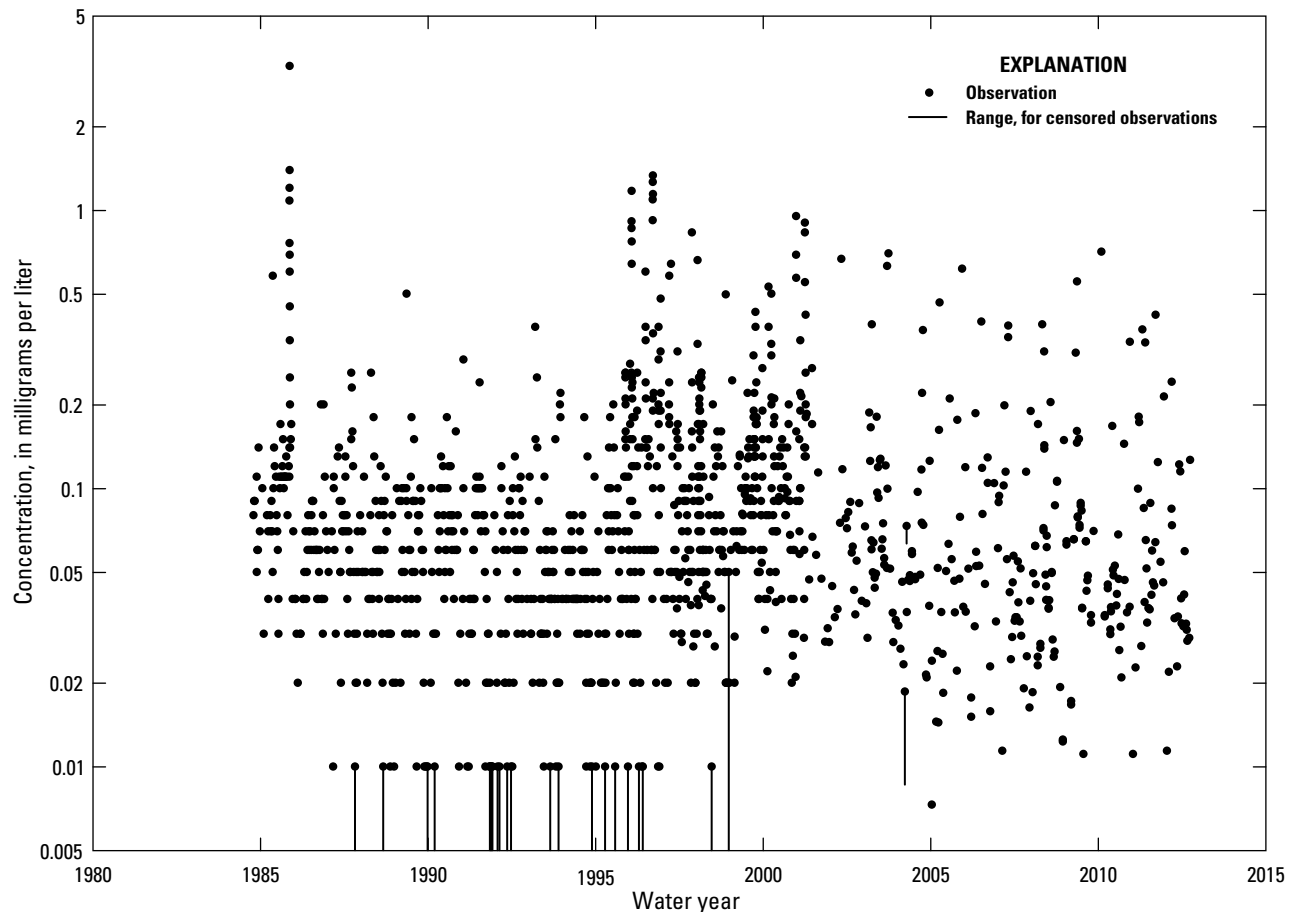


Figure 8. Total phosphorus concentration observations in relation to date at the Potomac River at Chain Bridge, at Washington, DC, for water years 1985–2012.

of WRTDS (that is, equation 4, with coefficients determined using weighted regression, as is currently done in WRTDS). The example that follows uses the long-term record for total phosphorus concentration from the Potomac River at Chain Bridge, at Washington, DC (POTC, USGS station 01646580; fig. 8). There is some evidence of a change in reporting convention, presumably associated with a change in laboratory methods around 1999, and considerable variability along the y-axis throughout the water-quality record for this station. Nevertheless, coarse visual examination suggests an overall downward trend in concentration. The flow-adjusted trend in concentration fitted by ESTIMATOR (fig. 9) is consistent with this visual impression, indicating a downward-parabolic trend, steepening toward the end of the record, with a period-of-record decrease of just under 20 percent.

The estimated concentration surface corresponding to figure 8, fit using weighted regression by WRTDS, is shown in figure 10.4. The surface is represented through color-coding: the red and gray shades correspond to the highest and lowest estimated concentrations, respectively; with the blue shades corresponding to intermediate concentrations. The most evident features are a pattern of increasing concentration

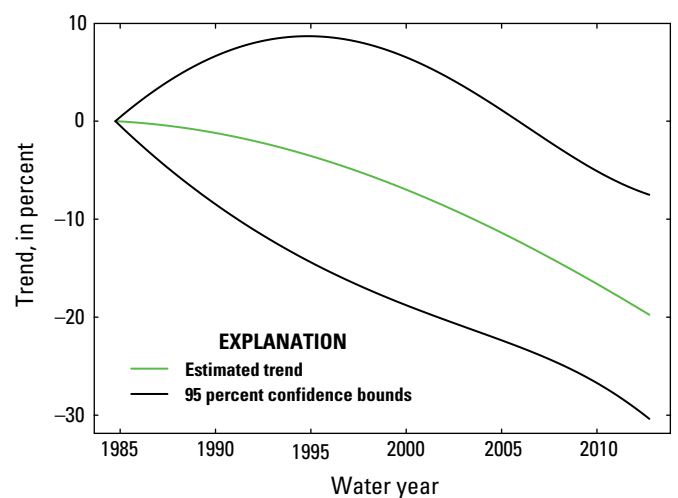
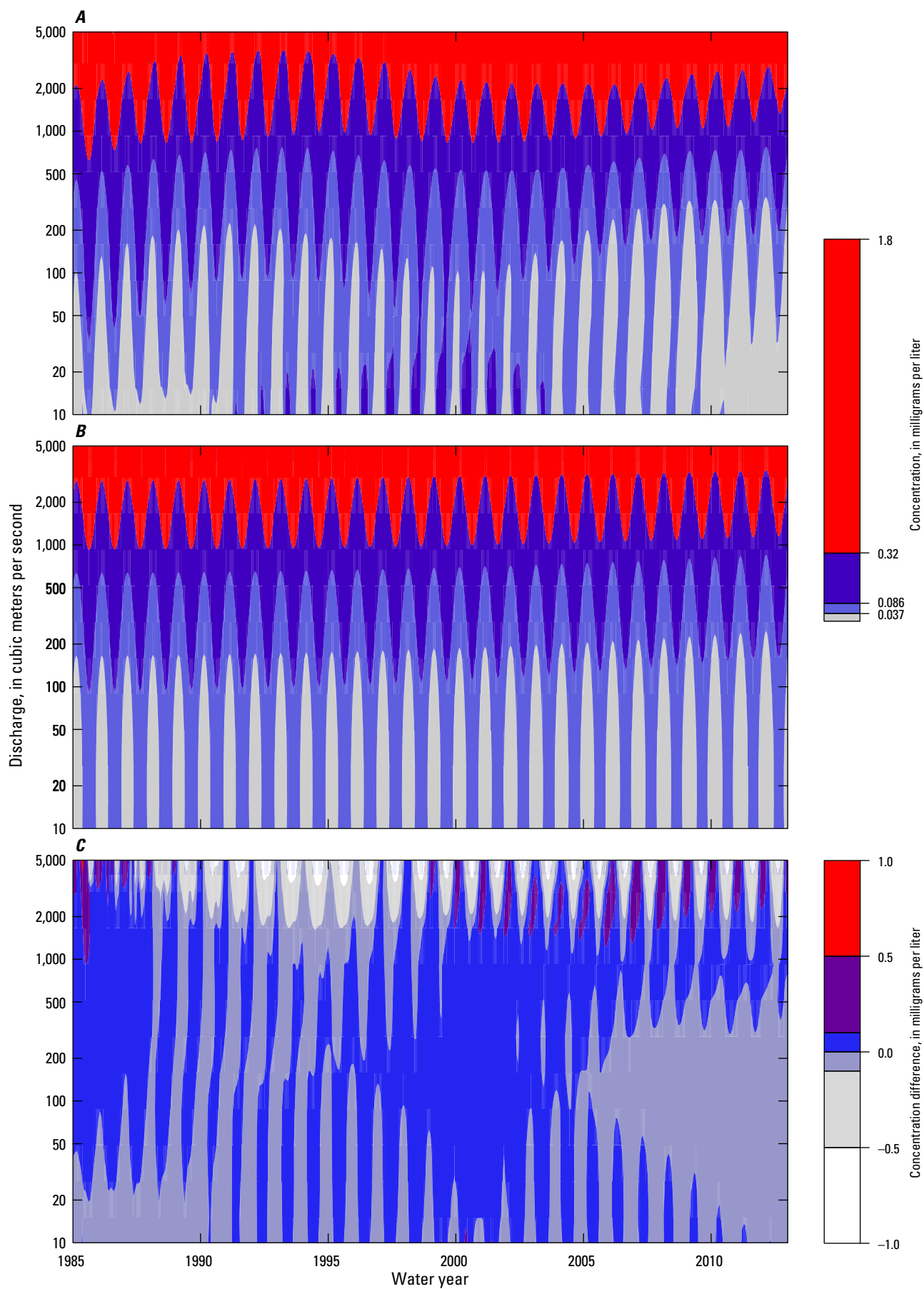


Figure 9. Flow-adjusted trend in concentration of total phosphorus at Potomac River at Chain Bridge, at Washington, DC, for water years 1985–2012 as computed by the ESTIMATOR model applied to the data shown in figure 8.



with increasing discharge, indicated by a gradient from gray to red upward along the y-axis, and a strong pattern of seasonality, indicated by the prominent lobes that occur in all concentration ranges with a period of 1 year. Somewhat less evident is the tendency for estimated concentration at any given discharge to be lower at the end of the record than at the beginning; this is perhaps most readily seen in the highest (red) and lowest (gray) ranges. For example, for a discharge range between 500 and 1,000 cubic meters per second (m^3/s), estimated concentrations exceed 0.32 mg/L (seasonally, red) at the beginning of the record, but not at the end. For a discharge range between 200 and 500 m^3/s , estimated concentrations never dip below 0.037 mg/L (gray) at the beginning of the record, but they do (seasonally) at the end. This overall downward trend in estimated concentrations is consistent both with the visual impression from the observed data and with the flow-adjusted trend fit by ESTIMATOR (fig. 9). Note also that the somewhat nuanced undulations evident over time in each concentration range can occur independently of one another. For example, for the discharge range between 10 and 20 m^3/s , estimated concentrations only exceeded 0.086 mg/L (seasonally, dark blue) in the middle of the record, most prominently around the year 2000; this pattern of timing is not evident in the other concentration ranges.

The concentration surface resulting from fitting the ESTIMATOR model to the same water-quality record, at the same grid points in the time-discharge domain shown for WRTDS in figure 10A, is shown in figure 10B. The major features, including the pattern of increasing concentration with increasing discharge and strong seasonality, appear similar to the WRTDS results discussed previously (fig. 10A). Also evident is the tendency for estimated concentrations at any chosen discharge and season to decrease over the period of record. The difference between the two figures lies in the regularity of figure 10B. ESTIMATOR is constrained to fit the entire dataset with linear/quadratic terms for both time and discharge, with no interaction terms. Thus, the surface shown in figure 10B is produced by translating an unvarying fitted concentration-discharge relation, defined by the linear and quadratic discharge coefficients, uniformly up and down the y-axis along a time trajectory defined by the linear and

quadratic time coefficients. A third surface, representing the difference between the concentration surfaces estimated by WRTDS and ESTIMATOR, can be used to identify regions of the time-discharge domain where WRTDS tends to predict a higher or lower concentration than ESTIMATOR (fig. 10C). This surface can be helpful in understanding differences in estimated trends between the two models, as discussed below.

The flow-normalized concentration on any day t of the record is the mean of the values interpolated from the concentration surface on that day (that is, a single location on the x-axes in figure 10) at all the discharges that occurred on that calendar day-of-year over the entire record (that is, multiple locations along the y-axis). Excepting a unit conversion factor, the corresponding flow-normalized flux is the mean of the product of the concentrations interpolated above and the discharges at which they were interpolated. Symbolically,

$$FNC(t) = \frac{1}{n} \sum_{i=1}^n C(t, Q_{T_i}) \quad (8)$$

and

$$FNF(t) = \frac{1}{n} \sum_{i=1}^n C(t, Q_{T_i}) \cdot Q_{T_i} \quad (9)$$

where

t	is a single day at any point in the record (T),
$FNC(t)$	is the flow-normalized concentration on day t (M/L^3),
$FNF(t)$	is the flow-normalized flux on day t (M/T),
n	is the number of years in the record,
$\{T_i\}, i=1, \dots, n$	is the set of all days in the record, including t , having the same calendar date (for example, "June 1") as t ,
$\{Q_{T_i}\}, i=1, \dots, n$	is the set of daily discharges observed on days T_i (L^3/T), and
$C(t, Q_{T_i})$	is the estimated concentration on day t and discharge Q_{T_i} (M/L^3).

Figure 10. (facing page) Estimated concentration surfaces as functions of time (x-axis) and discharge (y-axis) for total phosphorus at the Potomac River at Chain Bridge, at Washington, DC, based on the data shown in figure 8. A, Surface computed by weighted regression within WRTDS (Weighted Regressions on Time, Discharge, and Season). B, Surface computed by applying ESTIMATOR over the same domain. C, Arithmetic difference of A minus B.

Note that for any calendar day, flow-normalized values over sequential years are calculated using the same set of daily discharges—the set of all discharges observed on that calendar day. It is the concentrations at those discharges that change from year to year. Flow-normalized values at aggregated time scales, such as monthly or annual, are simply the mean of the daily flow-normalized values over the same period. Coarse analysis of long-term trends often considers annual values, because this time scale averages out seasonal variability. In the case of annual values, flow-normalized concentrations in any year also constitute averages over a repeating set of daily discharges each year—namely, all daily discharges occurring over the period of record, collapsed to lie vertically along the discharge axis of the time-discharge domain on their corresponding calendar days in any single year.

The time series of annual flow-normalized concentrations of total phosphorus at the Potomac River at Chain Bridge, derived from the WRTDS-estimated concentration surface (fig. 10A), is shown as a black line in figure 11A. The corresponding time series derived from applying the same flow-normalization process to the concentration surface estimated by ESTIMATOR (fig. 10B) is shown as a green line in figure 11A. The difference between the lines reflects the constrained evolution of the concentration-discharge relation over time for ESTIMATOR, relative to WRTDS. Note that periods where the annual flow-normalized concentrations reported using the WRTDS-derived concentration surface exceed those reported using the ESTIMATOR-derived concentration surface correspond to portions of the original surfaces where concentrations estimated by WRTDS exceed those estimated by ESTIMATOR over most of the discharge range, and vice versa (fig. 10C). The correspondence between figures 10C and 11A derives from the fact that concentrations in the most-frequently occurring discharges are those with the greatest influence on the average that flow-normalized concentration represents. The same lines shown in figure 11A, now expressed as a percentage difference in flow-normalized concentration relative to the value estimated by the corresponding model at the beginning of the record, are shown in figure 11B. Also shown in figure 11B is the relative flow-adjusted trend in concentration derived analytically using the ESTIMATOR model coefficients—this is the trend product historically reported for the RIM and CBNTN stations. That trend coincides with the trend resulting from applying the WRTDS flow-normalization procedure using the ESTIMATOR-derived concentration surface. Note also that the position of the endpoint of the WRTDS trend relative to the endpoints of the ESTIMATOR confidence interval in figure 11B matches the relative positions of the two models' estimates of period-of-record percentage trends for total phosphorus at the Potomac River at Chain Bridge station, summarized graphically with other long-term stations in figure 6B.

Recall that the ESTIMATOR flow-adjusted trend in concentration could be succinctly defined as “the trend in the observed concentration data that cannot be explained by

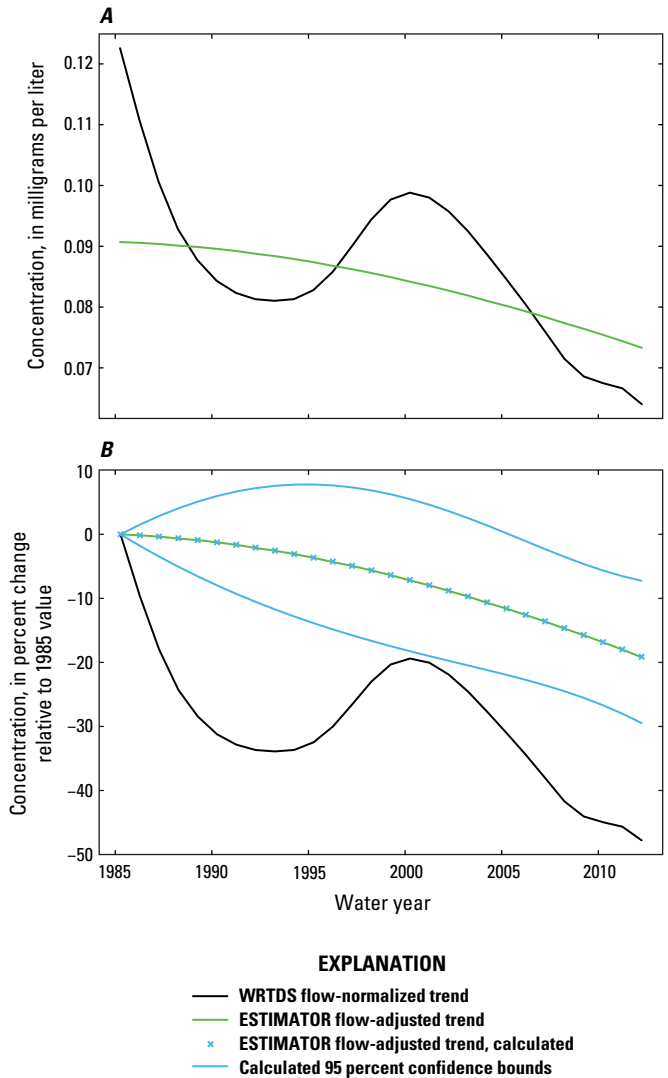


Figure 11. Period-of-record trends in concentration of total phosphorus at Potomac River at Chain Bridge, at Washington, DC, comparing flow-normalized trend computed using WRTDS (Weighted Regressions on Time, Discharge, and Season) with flow-adjusted trend computed using ESTIMATOR. *A*, Trends in concentration units. *B*, Trends as a percentage difference relative to water year 1985, with trend calculated analytically using ESTIMATOR model coefficients shown in blue.

discharge or season.” Although the flow-normalized trend in concentration produced by WRTDS is conceptualized and computed differently, trends shown in figure 11B indicate that, operationally, the same definition applies. This explains the broad pattern of similarity in the directions of the trends estimated by the two models in figure 6. Moyer and others (2012) and Hirsch (2014) cite a number of examples where the less-constrained approach to characterizing concentration variability with discharge used in WRTDS conveyed an apparent advantage, relative to ESTIMATOR, for flux

estimation. A comparison of long-term flow-normalized trends for total phosphorus at the nine RIM stations with corresponding ESTIMATOR flow-adjusted trends (re-scaled from relative to absolute concentration units; fig. 12) suggests that increased flexibility in WRTDS' representation of variability in concentration over both discharge and time conveys a similar advantage for trend reporting. Note that for shorter 10-year records (fig. 7), the two models produce much more comparable results for all three constituents considered. Recalling that the default half-window width for WRTDS time-weighting, that is, the distance from the point in time at which a concentration estimate is required to the point before or after at which the weighting drops from 1.0 to 0.0, is 10 years, the similarity of the 10-year trends in figure 7 likely reflects the fact that both WRTDS and ESTIMATOR are considering the same data, with WRTDS applying only a modest degree of time-weighting.

Trends in Flux

As discussed previously, the ESTIMATOR model, through the fitted coefficients on the time and time-squared terms in equation 1, can estimate a trend in concentration that cannot be explained by discharge or season. The equivalent trend for flux, expressed as a percentage change relative to a starting value, is constrained in ESTIMATOR to be the same as the trend in concentration. In terms of the model form itself, this can be appreciated by noting that, excepting unit conversion, the concentration model in equation 1 is transformed to a flux model by simply adding the logarithm of discharge to the right hand side. This adds 1 to the coefficient on linear discharge $\hat{\beta}_1$ and, through the discharge centering value, also affects the intercept $\hat{\beta}_0$, but neither change affects the calculation of $\hat{\tau}_{FA}$.

The constraint can also be appreciated in terms of the definitions of flow-normalization given in the previous section: Consider the same calendar day T in two sequential years, m and $m+1$. The change in flow-normalized concentration between those 2 days, expressed relative to the value in the first year is

$$\Delta FNC_{T_m, T_{m+1}} = 100 \cdot \frac{[FNC_{T_{m+1}} - FNC_{T_m}]}{FNC_{T_m}} \quad (10)$$

From equation 8 defining flow-normalized concentration,

$$\Delta FNC_{T_m, T_{m+1}} = 100 \cdot \frac{\left[\frac{1}{n} \sum_{i=1}^n C(T_{m+1}, Q_{T_i}) - \frac{1}{n} \sum_{i=1}^n C(T_m, Q_{T_i}) \right]}{\frac{1}{n} \sum_{i=1}^n C(T_m, Q_{T_i})} \quad (11)$$

Over a period of 1 year, seasonal effects cancel. Thus for the ESTIMATOR model, the change in concentration between 2 days separated by a year is

$$\begin{aligned} \Delta FNC_{T_m, T_{m+1}} &= 100 \cdot \frac{\left[\frac{1}{n} \sum_{i=1}^n e^{\hat{\tau}_{FA, 1yr}} \cdot C(T_m, Q_{T_i}) - \frac{1}{n} \sum_{i=1}^n C(T_m, Q_{T_i}) \right]}{\frac{1}{n} \sum_{i=1}^n C(T_m, Q_{T_i})} \\ &= 100 \cdot \frac{\left[\frac{e^{\hat{\tau}_{FA, 1yr}}}{n} \sum_{i=1}^n C(T_m, Q_{T_i}) - \frac{1}{n} \sum_{i=1}^n C(T_m, Q_{T_i}) \right]}{\frac{1}{n} \sum_{i=1}^n C(T_m, Q_{T_i})} \\ &= 100 \cdot \left(e^{\hat{\tau}_{FA, 1yr}} - 1 \right) \end{aligned} \quad (12)$$

where $\hat{\tau}_{FA,1yr}$ is the change in log concentration over a period of a year, calculated using equation 3 above. Note that the factoring leading to the above identity was only possible because for ESTIMATOR the change in log concentration is the same at all discharges $\{Q_{T_i}\}$.

The percentage change in flow-normalized flux is

$$\begin{aligned}\Delta FNF_{T_m, T_{m+1}} &= 100 \cdot \frac{[FNF_{T_{m+1}} - FNF_{T_m}]}{FNF_{T_m}} \\ &= 100 \cdot \frac{\left[\frac{1}{n} \sum_{i=1}^n C(T_{m+1}, Q_{T_i}) \cdot Q_{T_i} - \frac{1}{n} \sum_{i=1}^n C(T_m, Q_{T_i}) \cdot Q_{T_i} \right]}{\frac{1}{n} \sum_{i=1}^n C(T_m, Q_{T_i}) \cdot Q_{T_i}}\end{aligned}\quad (13)$$

In general, and for WRTDS weighted regression in particular, the change in concentration over any period can be different at different discharge values. However, as stated previously for ESTIMATOR, the change is constrained to be the same for all discharges. In that case the change in flow-normalized flux is

$$\begin{aligned}\Delta FNF_{T_m, T_{m+1}} &= 100 \cdot \frac{\left[\frac{1}{n} \sum_{i=1}^n e^{\hat{\tau}_{FA,1yr}} \cdot C(T_m, Q_{T_i}) \cdot Q_{T_i} - \frac{1}{n} \sum_{i=1}^n C(T_m, Q_{T_i}) \cdot Q_{T_i} \right]}{\frac{1}{n} \sum_{i=1}^n C(T_m, Q_{T_i}) \cdot Q_{T_i}} \\ &= 100 \cdot \frac{\left[\frac{e^{\hat{\tau}_{FA,1yr}}}{n} \sum_{i=1}^n C(T_m, Q_{T_i}) \cdot Q_{T_i} - \frac{1}{n} \sum_{i=1}^n C(T_m, Q_{T_i}) \cdot Q_{T_i} \right]}{\frac{1}{n} \sum_{i=1}^n C(T_m, Q_{T_i}) \cdot Q_{T_i}}\end{aligned}\quad (14)$$

Because all the summations in this last equation are carried out over the same set of discharge values $\{Q_{T_i}\}$, the summations can be factored. So, for ESTIMATOR

$$\Delta FNF_{T_m, T_{m+1}} = 100 \cdot \left(e^{\hat{\tau}_{FA,1yr}} - 1 \right) = \Delta FNC_{T_m, T_{m+1}} \quad (15)$$

Because flow-normalized concentration (or trend) values at larger time scales are the mean (or sum) of their constituent daily values, the identities demonstrated previously apply to any aggregation of daily values (for example, monthly or annual) separated by 1 year.

Just as a trend in flow-normalized concentration can be conceptualized as “the trend in the observed concentration data that cannot be explained by discharge or season,” the trend in flow-normalized flux can be conceptualized as “the trend in the observed flux data that cannot be explained by discharge or season.” Moyer and others (2012) provide several environmentally relevant examples of situations where change in concentration and change in flux differ in direction.

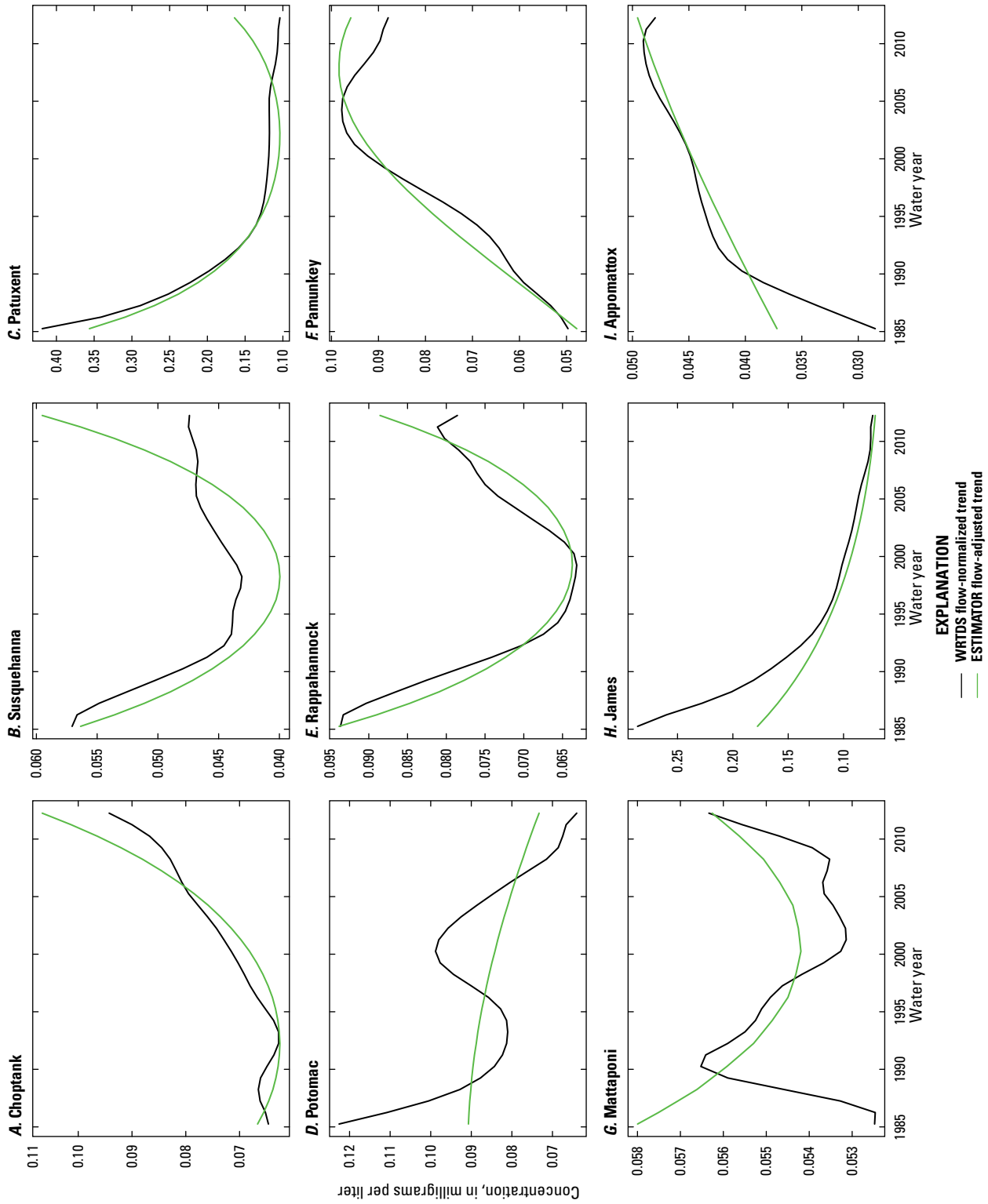


Figure 12. Period-of-record trends in concentration of total phosphorus at the nine Chesapeake Bay River Input Monitoring stations, comparing flow-normalized trends computed using WRTDS (Weighted Regressions on Time, Discharge, and Season) with flow-adjusted trends computed using ESTIMATOR.

Sensitivity of Trends to Incremental Incorporation of New Data

One of the primary uses of any water-quality trend result is to detect and interpret “turnarounds,” or inflection points, in concentration or flux. Because of the current regulatory environment in the Chesapeake Bay watershed, the most common motivation throughout the CBNTN is to credibly identify instances where point-source controls and land management actions are having a positive impact on riverine water quality. Regression-based trend estimation methods such as flow-adjustment and flow-normalization can be broadly thought of as techniques for smoothing a cloud of discrete observations into a refined product, often an annual time series, indicating trend over time with variability due to discharge and season removed. All smoothing methods that update their refined estimate as new data become available must somehow balance two competing desirable properties: for purposes of this discussion, they can be defined loosely as *sensitivity*, or the tendency to report a change as soon as it is indicated by the data, and *stability*, or the tendency for a newly reported change to accurately reflect the “true” situation as even more data are accumulated and the estimate “ages.” In light of the current priorities of the Bay Program, a comparison of ESTIMATOR and WRTDS should consider the influence of new water-quality data on both models’ trend estimates; that comparison

is the subject of this section. Briefly, the consequences of the tradeoff discussed above are found to be evident, to roughly the same degree, at the “recent” tail of both models’ reported trends, but for ESTIMATOR the effects of new data on the flow-adjusted trend can propagate throughout the trend estimate, in some cases resulting in appreciable differences in flow-adjusted concentration even at the beginning of the record. Because the sensitivity/stability tradeoff is an inherent limitation of all regression-based methods, caution is suggested in interpreting “turnarounds” that are less than several years old.

One straightforward means of examining the sensitivity of ESTIMATOR and WRTDS trend results to new data is to truncate a long water-quality record; build each model, treating the truncated data as complete; then iteratively add incremental “new” data and rebuild the models, treating each new dataset as the most “up-to-date.” This exercise was performed on the 1985–2012 record for total phosphorus at the Potomac River at Chain Bridge (fig. 8) beginning in 1994 and using water years as the incremental time scale. Water year 1994 was selected to fix the shortest resulting modeling period (1985–94) at 10 years, the default half-window width for the WRTDS model. The resulting sequence of estimated annual trend lines for flow-normalized concentration produced by WRTDS is shown in figure 13. (The longest trend in figure 13, spanning the years 1985–2012, is the same as the

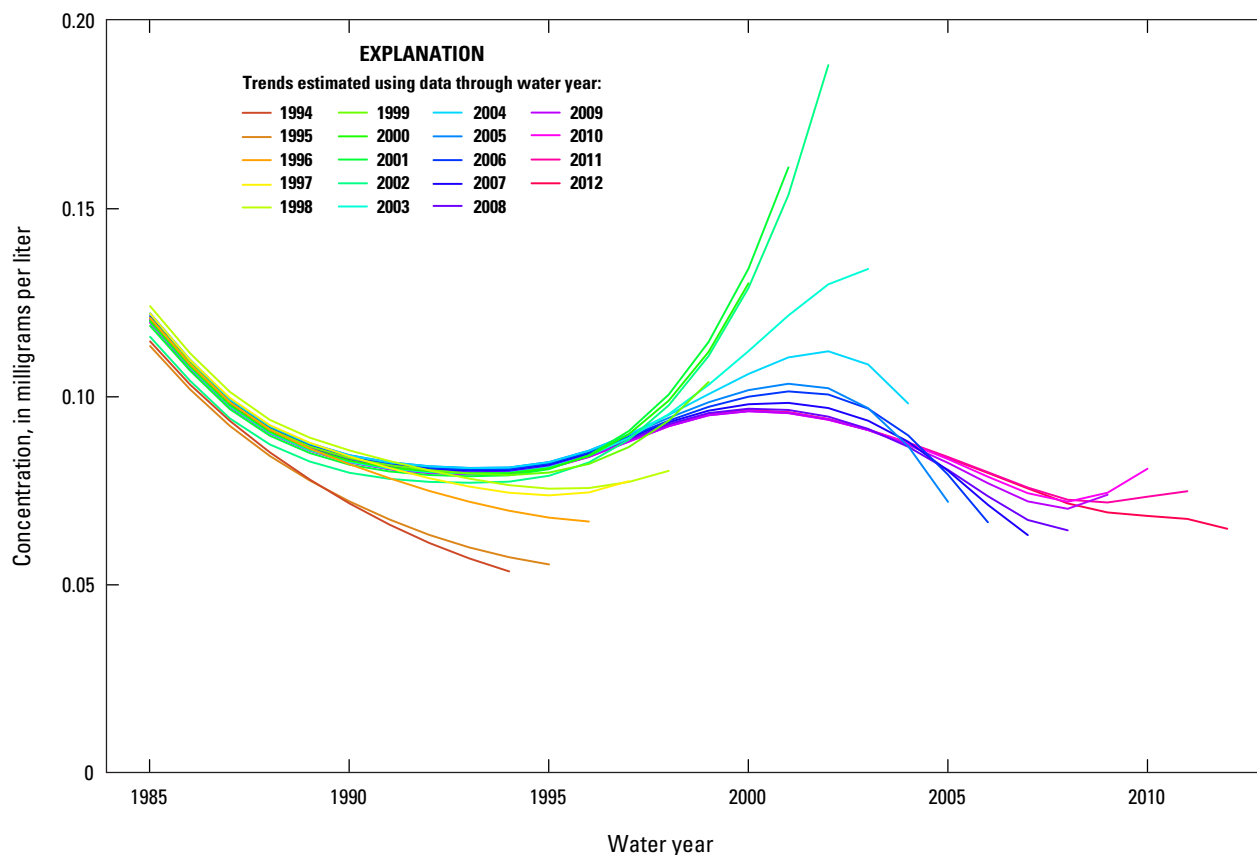


Figure 13. Trends in flow-normalized concentration of total phosphorus at Potomac River at Chain Bridge, at Washington, DC, estimated using WRTDS with records beginning in 1985 and ending in sequential years from 1994 through 2012.

trend shown in black in figure 11A). The small degree of up-and-down translation of the collection of trend lines in figure 13 is related to the flow-normalization procedure. Recall that for any annual estimate, the flow-normalization process for concentration averages estimated concentration values for all daily discharges observed over the period of record. Each year's new discharge values constitute an additional set of 365 ordered pairs per year, in the time-discharge domain, where concentration values interpolated from the estimated concentration surface are incorporated into the annual mean. Although the result is the degree of translation evident in figure 13, note that the *shape* of the trend in the early years changes very little: because of the 10-year half-window width and within-window time-weighting used by WRTDS, data added in later years have little or no effect on estimated values of the concentration surface at the beginning of the record.

Somewhat more problematic features evident in figure 13, in terms of their implications for interpretation, are the occasional instances of pronounced difference between point estimates of annual flow-normalized concentration in years that are at or near the presumptive end of the record and the estimated values for the same years based on the complete 1985–2012 record. For example, the two green lines near the top of the figure show the estimated annual flow-normalized trends, that is, the sequence of estimated annual flow-normalized concentration estimates, ending in water

years 2001 and 2002. Both suggest rather dramatic increases in flow-normalized concentration relative to, for example, year 2000. However, viewed retrospectively, the full dataset indicate that this period was actually an inflection point in the trend and suggest that the actual trend from 2000 through 2002 was slightly downward. The shape of the trends at the recent ends of the shortest records examined (1994–96, shown as red and orange lines in figure 13) hint that the same phenomenon, only reversed in sign, may have taken place in the early to mid-1990s; that is, initially, the data indicated a downward trend, but as more data were incorporated into the model the trend over the same period appears in retrospect to be slightly upward.

One reasonable question raised by the features in figure 13 is whether the behavior is somehow related to the time-windowing, time-weighting approach implemented in WRTDS and whether the un-weighted, constant-parameter model implemented in ESTIMATOR to yield flow-adjusted concentration might represent an improvement. Figure 14 shows the sequence of trends in flow-adjusted concentration, translated from relative percentage to absolute concentration units, produced by fitting the ESTIMATOR model to the same data used to produce figure 13. Consistent with the representation of time trends in ESTIMATOR by constant-parameter linear and quadratic time terms, flow-adjusted trends are constrained to evolve as a sequence of linear or

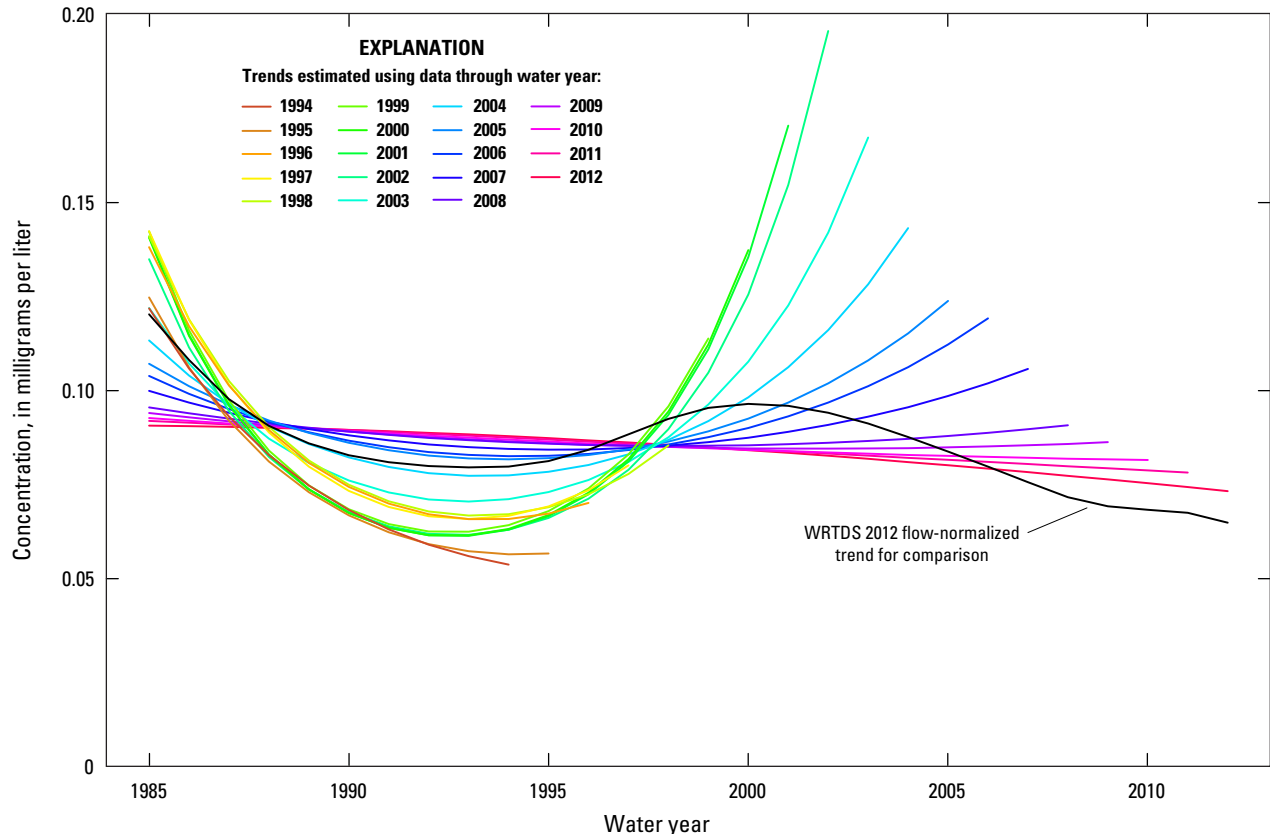


Figure 14. Trends in flow-adjusted concentration of total phosphorus at Potomac River at Chain Bridge, at Washington, DC, estimated using ESTIMATOR with records beginning in 1985 and ending in sequential years from 1994 through 2012. [WRTDS, Weighted Regressions on Time, Discharge, and Season]

quadratic curves, in this particular case reflecting a continuum from upward parabolic for the shortest modeling periods to nearly linear when fit to the period of record. (The longest trend in figure 14—the same as the trend shown in green in figure 11A—is contrasted in figure 14 with the corresponding period-of-record WRTDS trend, shown as a dotted black line.) The locations along the x-axis of the largest distances between flow-normalized concentration values estimated at the presumptive end of the record and their values based on the entire 1985–2012 record more-or-less correspond with those shown for WRTDS in figure 13, and it is not immediately evident that newly estimated flow-adjusted concentrations at those points vary any less from their values estimated using the entire record. Moreover, the “warping” of the flow-adjusted trends produced by ESTIMATOR influences flow-adjusted concentration estimates throughout the record, resulting in estimates for water year 1985 that wander in the range of 0.09 to 0.14 mg/L as annual data are incrementally added. This range of variability is greater than half of the estimate for water 1985 using data through 2012 (0.09 mg/L; fig. 14). To the extent that long-term records such as these are used to infer a “period-of-record” trend, that is, a change in either absolute concentration units or percentage relative to the beginning of the record, the additional degree of year-to-year variability in estimated flow-adjusted concentration at the beginning of the record confounds interpretation in a manner that lacks a plausible physical explanation. In contrast, the uniform up-and-down shifting of flow-normalized concentration trends as a new years’ daily discharge data are added in WRTDS (fig. 13; note in particular the early years of the record) could be expected to have a smaller effect on estimates of period-of-record change, because the phenomenon affects the entire record equally. Time series plots corresponding to figures 13 and 14 for two other RIM stations for which total phosphorus trends have multiple inflection points, the Rappahannock River near Fredericksburg, VA (RPHF, USGS station 01668000), and the Susquehanna River at Conowingo, MD (SRCW, USGS station 01578310), are included in appendix 1. To provide insight into how this phenomenon appears for a dissolved constituent, appendix 1 also includes plots for dissolved inorganic nitrogen at three stations for which trends for that constituent feature multiple inflection points.

The mechanism for the occasional large differences between flow-normalized values estimated at the presumptive end of the record and their retrospective values after additional years’ data have been collected and the features of the discharge and (or) water-quality record that result in these differences are topics of active ongoing investigation. Until the issue is better understood, it may be of value to quantify the effect of this phenomenon more comprehensively, because it is evidently a feature of both regression-based trend methods commonly used throughout the CBNTN. This quantification was done by tracking the evolution of estimates of flow-normalized concentration for a given year from the first year the estimate becomes available; that is, when the year for which the estimate is made is the presumptive end of the record, to some future point, beyond which the incorporation

of new data appears to have little retrospective influence on estimated flow-normalized concentration for the year that initially constituted the end of the record. For brevity, the latter estimate will be referred to as the “stable” estimate; trial-and-error experimentation on several stations and constituents indicated that estimates changed very little after 8–10 years. The procedure for a single station is illustrated graphically for total phosphorus at the Potomac River at Chain Bridge in figure 15. Four “comparison” water years—1994, 1997, 2000, and 2003—were chosen for recording estimates of flow-normalized concentration. Simulations were then run for the set of periods 1985–94 through 1985–2012. As the simulation end year expanded to include each comparison year, the flow-normalized concentration for that year was recorded. The value in that comparison year was then recorded for simulations including from 1 to 9 subsequent water years; the value in the comparison year after nine additional years’ data were included was recorded as the stable value for the comparison year (fig. 15A). The value recorded in each comparison year and the nine subsequent years was expressed as a percentage difference from the stable value, expressed as $(\text{“value } i \text{ years after comparison year”} - \text{stable value}) \times 100 / \text{stable value}$, with i ranging from 0 to 9. Note that by definition, percentage differences from the stable value decay to 0 after 9 years. With four “comparison” water years defined in this manner, this scheme permitted the construction of four time series of percentage differences, as shown in figure 15B. Taken together, these four time series indicate that, for example, annual estimates of flow-normalized total phosphorus concentration at this station that are about 4–5 years old are within about 10 percent of their stable value, as defined above. The procedure illustrated for total phosphorus at the Potomac River at Chain Bridge, shown in figure 15, was then repeated for the same period of record at two other RIM stations: Rappahannock River near Fredericksburg, VA (RPHF, USGS station 01668000), and Susquehanna River at Conowingo, MD (SRCW, USGS station 01578310). The three stations were selected because their flow-normalized concentration time series for total phosphorus were among the “busiest” of the nine RIM stations. This combination of three different stations and four “comparison” years, each spaced 3 years apart, was taken as a set of 12 “pseudo-independent” observations of trend stability; pooled variability as a function of years since the trend was first estimated was represented using boxplots with $n=12$ observations used to construct each box. The experiment as described for WRTDS was then repeated using ESTIMATOR, with flow-adjusted concentrations in percentage relative to 1985 re-cast into absolute concentration units. To compare trend stability for a dissolved constituent, the entire experiment was then repeated using 1985–2012 records of dissolved inorganic nitrogen at three other RIM stations: James River at Cartersville, VA (JAMC, USGS station 02035000), Mattaponi River near Beulahville, VA (MABU, USGS station 01674500), and Pamunkey River near Hanover, VA (PAMU, USGS station 01673000).

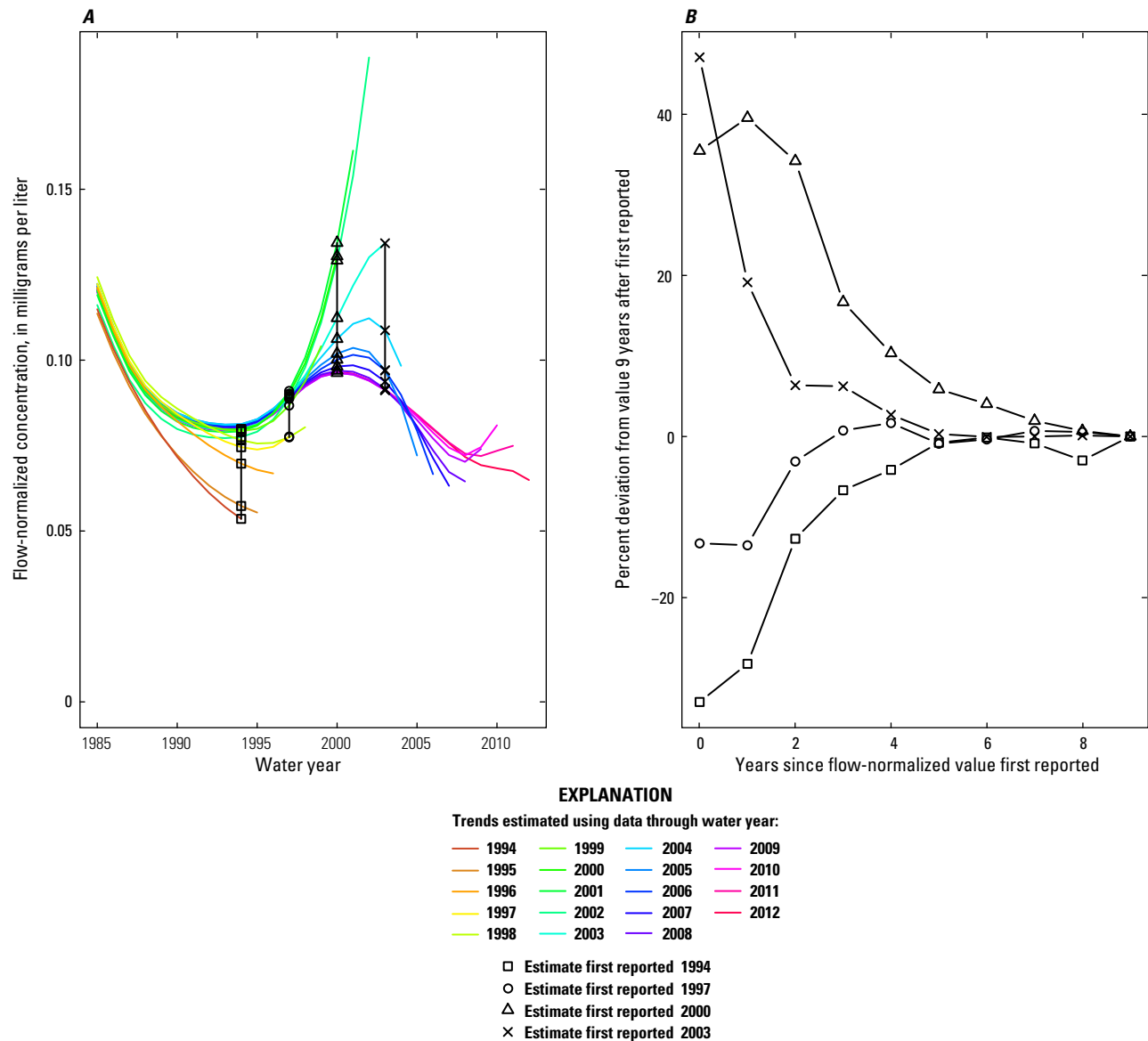


Figure 15. The evolution in estimated values of flow-normalized concentration of total phosphorus at Potomac River at Chain Bridge, at Washington, DC, for the years 1994, 1997, 2000, and 2003. *A*, The sequential trends shown in figure 13, with vertical black lines depicting the change in estimated flow-normalized concentrations as incremental years of data are added. *B*, Time-series plots showing the percentage deviation of the estimated flow-normalized concentration in 1994, 1997, 2000, and 2003 from the values estimated with 9 additional years' retrospective data.

For both constituents, ESTIMATOR and WRTDS showed similar patterns and ranges of variability in estimated annual flow-normalized/flow-adjusted concentration as additional years' data became available, although the magnitude of variability was considerably larger for total phosphorus (fig. 16) than for dissolved inorganic nitrogen (fig. 17; note smaller y-axis range). For total phosphorus, about half of the values estimated by either model in the first year were within plus or minus 20 percent of their stable value. For WRTDS, all but one of the 12 estimates were within 20 percent of their stable value after 2 additional years' data were collected (that

is, the third year the estimate was reported), and all values were within 10 percent of their stable values after 4 additional years' data were collected. For ESTIMATOR, all values were within 10 percent of their stable value after 5 additional years' data were accumulated. With regard to the median percentage differences, the systematic tendency for both models to underestimate the stable value in the first year and overestimate in the second and subsequent years before converging on the stable value is probably unique to the specific set of values encountered in this rather small sample size and should not be taken as representative of expected behavior at all CBNTN

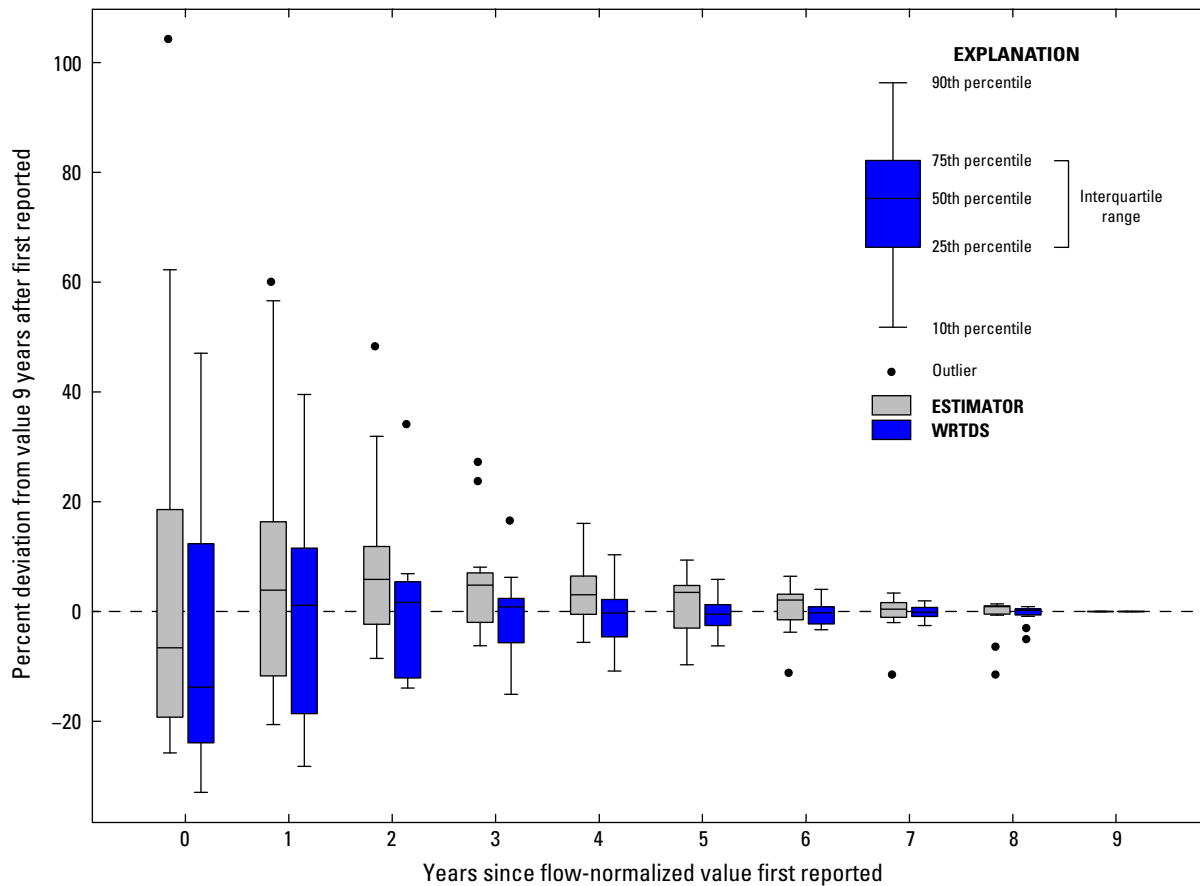


Figure 16. Boxplots showing the deviation of both WRTDS (Weighted Regressions on Time, Discharge, and Season) and ESTIMATOR estimates of flow-normalized concentration of total phosphorus for 1994, 1997, 2000, and 2003 at three Chesapeake Bay Nontidal Water-Quality Monitoring Network (CBNTN) stations from the values estimated with retrospective data for 9 additional years.

stations. For dissolved inorganic nitrogen, nearly all the values estimated by either model were within plus or minus 20 percent of their stable value the first year they were reported, and all values for either model were within 10 percent of their stable value after 3 additional years' data were collected.

Because the stations for this experiment were deliberately selected to represent the “busier” trends among the RIM stations, the boxplots in figures 16 and 17 probably indicate the more extreme range of the degree of variability associated with incorporation of new data that could be expected network-wide. However, variability of this sort exists independently of uncertainty due to sampling variability; thus, variability associated with incorporating new data can be expected to be additive to the variability associated with

sampling-associated confidence intervals currently in existence for ESTIMATOR and under development for WRTDS. The issue of exaggerated curvature near the beginning and end of the record has been recognized in several WRTDS applications within and outside of the Chesapeake Bay watershed. The default method in the EGRET software (version 2.0) implementation of WRTDS introduces a modification of the computations, called edgeAdjust, (Hirsch and De Cicco, 2014), to reduce the extent of this type of curvature; the use of this modification in future applications to the CBNTN is under consideration. It should be recognized that this type of variability likely reflects the unavoidable tradeoff between sensitivity and stability, as discussed previously, inherent in all regression-based flux and trend methods.

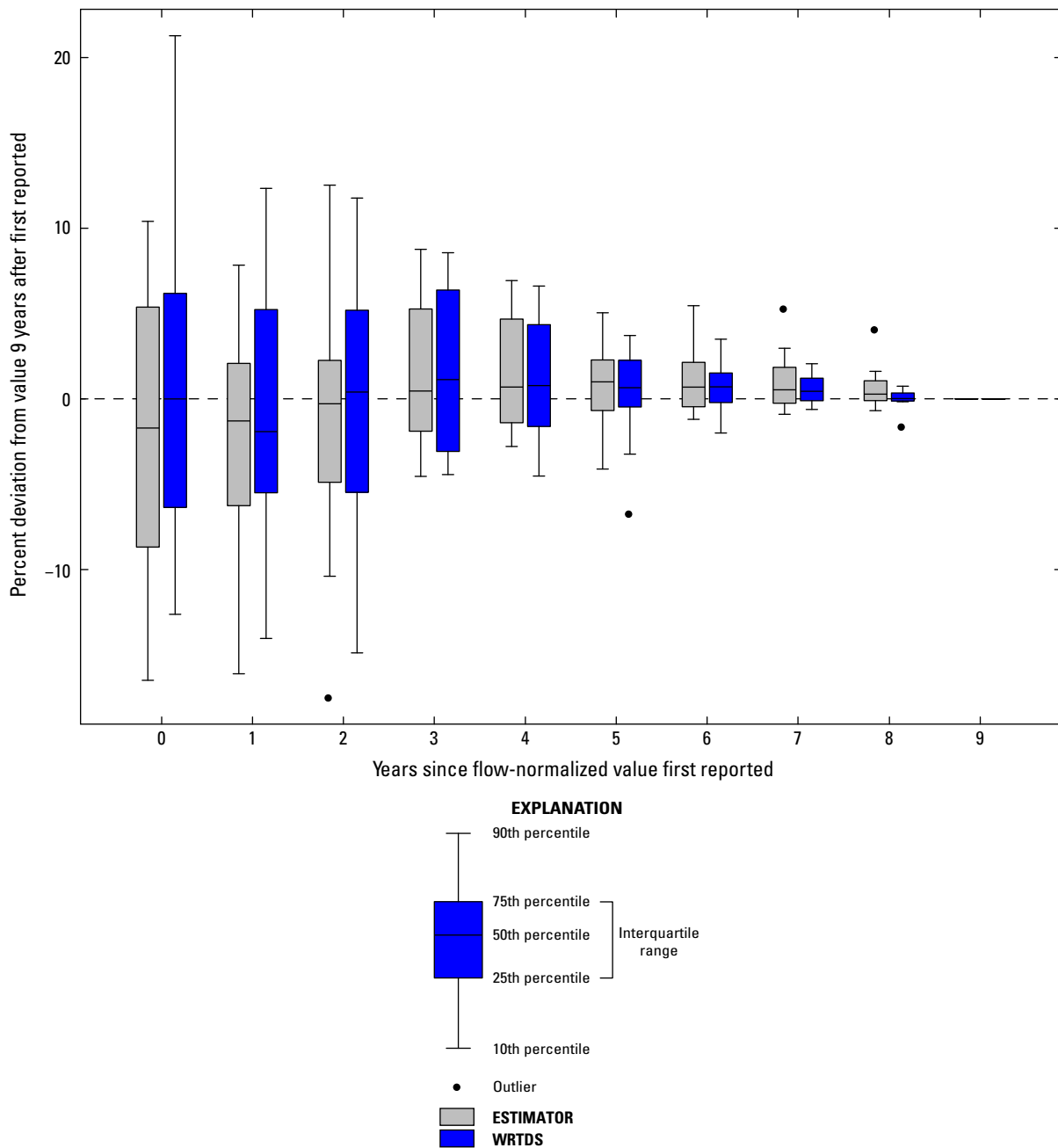


Figure 17. Boxplots showing the deviation of both WRTDS (Weighted Regressions on Time, Discharge, and Season) and ESTIMATOR estimates of flow-normalized concentration of dissolved inorganic nitrogen for 1994, 1997, 2000, and 2003 at three Chesapeake Bay Nontidal Water-Quality Monitoring Network (CBNTN) stations from the values estimated with retrospective data for 9 additional years.

Summary of Model Comparison Findings

The comparisons made in this section of the report can be summarized as follows:

1. Application of WRTDS to estimate fluxes across a wide range of records from the CBNTN demonstrated that the same advantages over ESTIMATOR reported by Moyer and others (2012) and Hirsch (2014), who used carefully selected water-quality records, are evident in the broader setting of an operational water-quality network.
2. WRTDS estimates of flow-normalized trend in concentration generally tracked ESTIMATOR flow-adjusted trend in concentration for stations having record lengths exceeding 25 years; the correspondence became much closer for stations with 10-year records.
3. In light of the operational similarities in the models' trend estimation approaches, despite conceptual differences, the flow-normalization approach used in WRTDS can be thought of as a more flexible extension of that used by ESTIMATOR, rather than a new, unrelated approach.
4. By virtue of that flexibility, WRTDS was demonstrated to produce more credible estimates of long-term concentration trends at the nine RIM stations than ESTIMATOR.
5. Whereas ESTIMATOR is constrained to reporting trends in concentration only, the model structure of WRTDS allows for independent estimation of trends in concentration and flux; because trends in fluxes and trends in concentration reflect the effects of different watershed processes, the ability to estimate trends in fluxes represents an advantage for interpretation.
6. Trends estimated by both ESTIMATOR and WRTDS showed a similar degree of sensitivity to incremental addition of new data.

In all, these findings indicate WRTDS has several significant advantages, for both flux and trend estimation over ESTIMATOR, and no significant disadvantages. On that basis, WRTDS is adopted as the primary tool for future reporting of both flux and trend results for the CBNTN effective on the next reporting cycle, currently projected to include data through water year 2014.

WRTDS Estimated Flux and Trend Results for Nutrients and Sediment for the Chesapeake Bay Nontidal Water-Quality Monitoring Network Through Water Year 2012

On the basis of the recommendation made in the previous section, the five "internal" scenarios used to produce results for the CBNTN through water year 2012, outlined in the section "Description of Scenarios Considered," were re-visited

in order to assemble flux and trend results using WRTDS exclusively. For the *LONG_TERM* and *10_YEAR* scenarios, WRTDS had already been run for the purpose of reporting fluxes and yields; WRTDS concentration and flux trend results reported herein were simply drawn from the output of those model runs. The *10_YEAR_SECONDARY* scenario was originally run using ESTIMATOR for the purpose of determining trends; those scenarios were re-run using WRTDS to generate the results reported below. Fluxes originally reported from the *LESS_THAN_10_WRTDS* scenario are reported unchanged. On the basis of more recent experience with application of WRTDS to stations with as few as 60 observations (Hirsch and De Cicco, 2014), fluxes originally generated for stations with fewer than 120 observations under the *LESS_THAN_10_ESTIMATOR* scenario were re-run using WRTDS. Because in some cases the *LESS_THAN_10_ESTIMATOR* stations had fewer than the default minimum of 100 observations and 50 uncensored observations used in the EGRET package implementation of WRTDS for running a single instance of survival regression, those minima were set to 90 percent of the actual number of observations for each station-constituent combination run under the *LESS_THAN_10_ESTIMATOR* scenario.

The following subsections provide a tabular summary of WRTDS flux and yield results, including graphical summaries of WRTDS period-of-record and 10-year flow-normalized trends for total nitrogen, total phosphorus, and sediment. Tabular summary results for period-of-record and 10-year trends for total nitrogen, total phosphorus, and sediment are included in appendix 2. Full listings of WRTDS estimates of annual concentration, flux, flow-normalized concentration, and flow-normalized flux for total nitrogen, dissolved inorganic nitrogen, total phosphorus, orthophosphate, and sediment at all CBNTN stations modeled through water year 2012 are included in appendix 3. Data listed in appendix 3 should be considered the final record, through water year 2012, for purposes of management evaluation/decision support and for scientific interpretation.

Fluxes and Yields

The 5-year (2008–12) average annual yield for total nitrogen at the 65 CBNTN stations where total nitrogen was reported ranged from 117 kilograms per year per square kilometer ($\text{kg yr}^{-1} \text{ km}^{-2}$) to 3,460 $\text{kg yr}^{-1} \text{ km}^{-2}$, a range spanning nearly a factor of 30 (table 3). The network-wide 25th percentile, median, and 75th percentile yields were 406, 687, and 1,200 $\text{kg yr}^{-1} \text{ km}^{-2}$, respectively. Among the 9 RIM stations, which represent 78 percent of the freshwater inflow to the bay, the highest yields were reported at the Susquehanna River at Conowingo, MD (SRCW, USGS station 01578310), with a yield of 921 $\text{kg yr}^{-1} \text{ km}^{-2}$, and the Choptank River near Greensboro, MD (CHOP, USGS station 01491000), with a yield of 834 $\text{kg yr}^{-1} \text{ km}^{-2}$. The lowest-yielding RIM station was the Appomattox River at Matoaca, VA (APPM, USGS station 02041650), with a yield of 126 $\text{kg yr}^{-1} \text{ km}^{-2}$.

Table 3. Water year 2012 fluxes and 5-year (2008–12) mean yields for total nitrogen, total phosphorus, and sediment at the Chesapeake Bay Nontidal Water-Quality Monitoring Network stations with record lengths of at least 5 years.

[Stations are listed in downstream order by U.S. Geological Survey station identification number. kg yr⁻¹, kilograms per year; kg yr⁻¹ km⁻², kilograms per year per square kilometer. Sediment data type: TSS, total suspended solids; SSC, suspended-sediment concentration. Pink shading denotes River Input Monitoring (RIM) site]

Station abbreviation	Station ID	Station name	Total nitrogen		Total phosphorus		Sediment		Sediment data type
			2012 load (kg yr ⁻¹)	2008–2012 mean yield (kg yr ⁻¹ km ⁻²)	2012 load (kg yr ⁻¹)	2008–2012 mean yield (kg yr ⁻¹ km ⁻²)	2012 load (kg yr ⁻¹)	2008–2012 mean yield (kg yr ⁻¹ km ⁻²)	
NANT	01487000	Nanticoke River near Bridgeville, DE	287,000	2,080	2,520	37.5	236,000	4,800	TSS
MARS	01488500	Marshyhope Creek near Adamsville, DE	108,000	1,330	4,140	107	415,000	10,600	TSS
CHOP	01491000	Choptank River near Greensboro, MD	204,000	834	13,900	69.7	2,120,000	9,840	SSC
TUCK	01491500	Tuckahoe Creek near Ruthsburg, MD	324,000	1,600	9,800	70.1	1,180,000	8,200	TSS
BGEL	01495000	Big Elk Creek at Elk Mills, MD	274,000	2,070	10,700	132	7,890,000	179,000	TSS
UNAD	01502500	Unadilla River at Rockdale, NY	699,000	694	25,800	52.5	15,800,000	55,700	TSS
SRCK	01503000	Susquehanna River at Conklin, NY	1,830,000	529	104,000	51.1	59,000,000	41,700	TSS
SRWA	01515000	Susquehanna River near Waverly NY	5,080,000	616	328,000	62.7	181,000,000	59,900	SSC
COHO	01529500	Cohocton River near Campbell NY	590,000	599	15,900	27.7	6,350,000	14,000	TSS
CHEM	01531000	Chemung River at Chemung NY	2,300,000	452	197,000	48.7	133,000,000	48,700	SSC
SRTW	01531500	Susquehanna River at Towanda, PA	7,300,000	494	448,000	51.1	249,000,000	43,100	SSC
SRWB	01536500	Susquehanna River at Wilkes-Barre, PA	9,980,000	529	676,000	76.4	481,000,000	101,000	SSC
SRDA	01540500	Susquehanna River at Danville, PA	12,900,000	581	844,000	63.0	585,000,000	53,900	SSC
WBSK	01542500	West Branch Susquehanna River at Karthaus, PA	1,340,000	364	32,700	13.6	78,500,000	23,300	SSC
WBSJ	01549760	West Branch Susquehanna River near Jersey Shore, PA	5,400,000	368	145,000	16.1	136,000,000	14,700	SSC
WBSL	01553500	West Branch Susquehanna River at Lewisburg, PA	8,390,000	476	280,000	23.1	187,000,000	19,900	SSC
PENN	01555000	Penns Creek at Penns Creek, PA	798,000	956	20,300	41.7	14,800,000	31,600	SSC
RBJU	01562000	Raystown Branch Juniata River at Saxton, PA	1,720,000	911	44,500	35.0	38,700,000	42,700	SSC
JUNI	01567000	Juniata River at Newport, PA	6,890,000	816	217,000	33.4	104,000,000	19,500	SSC
SHER	01568000	Sherman Creek at Shermans Dale, PA	581,000	1,190	22,300	63.7	15,800,000	46,900	SSC
COND	01570000	Conodoguinet Creek near Hogestown, PA	2,650,000	2,080	41,300	41.3	45,800,000	42,700	SSC
YELL	01571500	Yellow Breeches Creek near Camp Hill, PA	776,000	1,310	18,600	51.5	8,660,000	40,600	SSC
SWAT	01573560	Swatara Creek near Hershey, PA	2,610,000	2,410	59,900	135	45,400,000	227,000	SSC
WCWG	01574000	West Conewago Creek near Manchester, PA	1,780,000	1,420	161,000	151	71,200,000	88,600	SSC
SRMA	01576000	Susquehanna River at Marietta, PA	46,700,000	785	2,170,000	53.6	1,400,000,000	43,100	SSC
CONE	01576754	Conestoga River at Conestoga, PA	3,960,000	3,460	145,000	153	37,500,000	50,400	SSC
PEQU	01576787	Pequea Creek near Martic Forge, PA	1,280,000	3,390	70,300	202	22,600,000	72,900	SSC
SRCW	01578310	Susquehanna River at Conowingo, MD	54,000,000	921	2,160,000	66.6	789,000,000	71,800	SSC
OCTO	01578475	Octoraro Creek at Richardsmere, MD	1,410,000	3,030	27,600	87.6	5,530,000	26,100	TSS
DEER	01580520	Deer Creek near Darlington, MD	721,000	1,720	18,100	66.9	13,900,000	77,400	TSS
GUNP	01582500	Gunpowder Falls at Glencoe, MD	553,000	1,190	10,300	23.9	9,030,000	23,800	TSS
NBPA	01586000	North Branch Patapsco River at Cedarhurst, MD	256,000	1,560	6,670	41.0	7,800,000	43,800	TSS
GWYN	01589300	Gwynns Falls at Villa Nova, MD	59,400	774	3,340	52.9	3,970,000	81,300	TSS
PAXU	01591000	Patuxent River near Unity, MD	113,000	1,100	3,550	32.5	4,670,000	46,200	TSS
PAXB	01594440	Patuxent River near Bowie, MD	494,000	609	38,900	55.7	18,000,000	28,200	SSC

40 Application of a Weighted Regression Model for the Chesapeake Bay Nontidal Water-Quality Monitoring Network

Table 3. Water year 2012 fluxes and 5-year (2008–12) mean yields for total nitrogen, total phosphorus, and sediment at the Chesapeake Bay Nontidal Water-Quality Monitoring Network stations with record lengths of at least 5 years.—Continued

[Stations are listed in downstream order by U.S. Geological Survey station identification number. kg yr⁻¹, kilograms per year; kg yr⁻¹ km⁻², kilograms per year per square kilometer. Sediment data type: TSS, total suspended solids; SSC, suspended-sediment concentration. Pink shading denotes River Input Monitoring (RIM) site]

Station abbreviation	Station ID	Station name	Total nitrogen		Total phosphorus		Sediment		Sediment data type
			2012 load (kg yr ⁻¹)	2008–2012 mean yield (kg yr ⁻¹ km ⁻²)	2012 load (kg yr ⁻¹)	2008–2012 mean yield (kg yr ⁻¹ km ⁻²)	2012 load (kg yr ⁻¹)	2008–2012 mean yield (kg yr ⁻¹ km ⁻²)	
WBUM	01594526	Western Branch at Upper Marlboro, MD	56,700	448	12,000	106	7,260,000	91,800	TSS
GEOR	01599000	Georges Creek at Franklin, MD	83,900	637	3,160	36.8	5,310,000	83,000	TSS
WILL	01601500	Wills Creek near Cumberland, MD	288,000	676	12,000	36.8	15,500,000	94,600	TSS
PATT	01604500	Patterson Creek near Headsville, WV	192,000	340	8,570	20.2	6,260,000	16,600	SSC
SBPO	01608500	South Branch Potomac River near Springfield, WV	980,000	304	76,700	25.1	30,300,000	16,800	SSC
TOWN	01609000	Town Creek near Oldtown, MD	103,000	337	2,330	13.7	8,030,000	77,400	SSC
SIDE	01610155	Sideling Hill Creek near Bellegrove, MD	106,000	406	2,200	12.0	4,630,000	33,100	TSS
CACA	01611500	Cacapon River near Great Cacapon, WV	440,000	305	24,900	22.7	17,200,000	29,700	SSC
TONO	01613095	Tonoloway Creek near Hancock, MD	195,000	725	3,780	14.4	3,930,000	19,000	TSS
LICK	01613525	Licking Creek near Pectonville, MD	364,000	687	14,500	28.4	16,100,000	41,300	TSS
CONC	01614500	Conococheague Creek at Fairview, MD	2,920,000	2,080	65,800	57.4	27,800,000	36,100	TSS
OPEQ	01616500	Opequon Creek near Martinsburg, WV	369,000	687	18,100	50.8	5,310,000	25,400	SSC
ANTW	01619000	Antietam Creek near Waynesboro, PA	494,000	1,760	16,800	68.0	8,440,000	33,900	TSS
ANTS	01619500	Antietam Creek near Sharpsburg, MD	1,510,000	1,720	42,700	60.2	14,700,000	22,200	TSS
SFSF	01631000	South Fork Shenandoah River at Front Royal, VA	1,560,000	406	86,600	43.1	34,200,000	28,800	SSC
NFSH	01634000	North Fork Shenandoah River near Strasburg, VA	925,000	490	35,100	41.7	22,700,000	40,600	SSC
CATM	01637500	Catoctin Creek near Middletown, MD	154,000	774	9,890	62.7	10,600,000	76,700	TSS
MONO	01639000	Monocacy River at Bridgeport, MD	553,000	1,200	59,000	124	23,500,000	49,400	TSS
POTC	01646580	Potomac River at Chain Bridge, at Washington, DC	17,100,000	676	934,000	50.1	531,000,000	46,600	SSC
NWBA	01651000	Northwest Branch Anacostia near Hyattsville, MD	64,900	725	5,760	92.1	6,120,000	136,000	SSC
RPDC	01667500	Rapidan River near Culpeper, VA	576,000	525	109,000	161	74,800,000	152,000	SSC
RPHF	01668000	Rappahannock River near Fredericksburg, VA	1,470,000	396	272,000	82.0	176,000,000	69,000	SSC
NANN	01671020	North Anna River at Hart Corner near Doswell, VA	160,000	117	16,200	12.5	9,120,000	8,090	TSS
PAMU	01673000	Pamunkey River near Hanover, VA	599,000	177	75,700	23.5	45,200,000	14,500	SSC
MABU	01674500	Mattaponi River near Beulahville, VA	267,000	165	25,400	15.9	5,490,000	3,920	SSC
JAMB	02024752	James River at Blue Ridge Parkway near Big Island, VA	1,680,000	202	198,000	29.6	149,000,000	24,200	SSC
JAMC	02035000	James River at Cartersville, VA	4,160,000	245	816,000	51.5	576,000,000	38,500	SSC
JAMR	02037500	James River near Richmond, VA	4,540,000	230	835,000	47.6	485,000,000	26,900	TSS
APPM	02041650	Appomattox River at Matoaca, VA	296,000	126	25,200	12.4	5,760,000	3,250	SSC
CHIC	02042500	Chickahominy River near Providence Forge, VA	108,000	214	15,100	32.9	1,900,000	6,970	TSS
Minimum			56,700	117	2,200	12.0	236,000	3,250	
25th percentile			274,000	406	12,000	29.6	6,350,000	22,200	
Median			699,000	687	27,600	50.8	16,100,000	40,600	
75th percentile			2,300,000	1,200	145,000	66.9	74,800,000	59,900	
Maximum			54,000,000	3,460	2,170,000	202	1,400,000,000	227,000	

The range of 5-year yields for total phosphorus spanned a factor of about 17, or about half that of total nitrogen, with minimum and maximum values of 12 and 202 kg yr⁻¹ km⁻², respectively (table 3). The median yield among the 65 stations was 50.8 kg yr⁻¹ km⁻², and the 25th and 75th percentiles were 29.6 and 66.9 kg yr⁻¹ km⁻², respectively. The highest-yielding RIM stations were the Rappahannock River near Fredericksburg, VA (RPHF, USGS station 01668000), with a yield of 82 kg yr⁻¹ km⁻², and the Choptank River near Greensboro, MD (CHOP, USGS station 01491000), with a yield of 69.7 kg yr⁻¹ km⁻². The lowest-yielding RIM stations were the Appomattox, Mattaponi, and Pamunkey Rivers in Virginia (APPM, MABU, and PAMU; USGS stations 02041650, 01674500, and 01673000, respectively), with yields of 12.4, 15.9, and 23.5 kg yr⁻¹ km⁻², respectively.

Sediment yields were considerably more variable than either total nitrogen or total phosphorus, with minimum, median, and maximum values of 3,250, 40,600, and 227,000 kg yr⁻¹ km⁻², respectively, a range spanning a factor of about 70 (table 3). The highest-yielding RIM stations were the Susquehanna River at Conowingo, MD (SRCW, USGS station 01578310), with a sediment yield of 71,800 kg yr⁻¹ km⁻², and the Rappahannock River near Fredericksburg, VA (RPHF, USGS station 01668000), with a yield of 69,000 kg yr⁻¹ km⁻². The Appomattox River near Matoaca, VA (APPM, USGS station 02041650), and the Mattaponi River near Beulahville, VA (MABU, USGS station 01674500), were the lowest-yielding RIM stations, with yields of 3,250 and 3,920 kg yr⁻¹ km⁻², respectively.

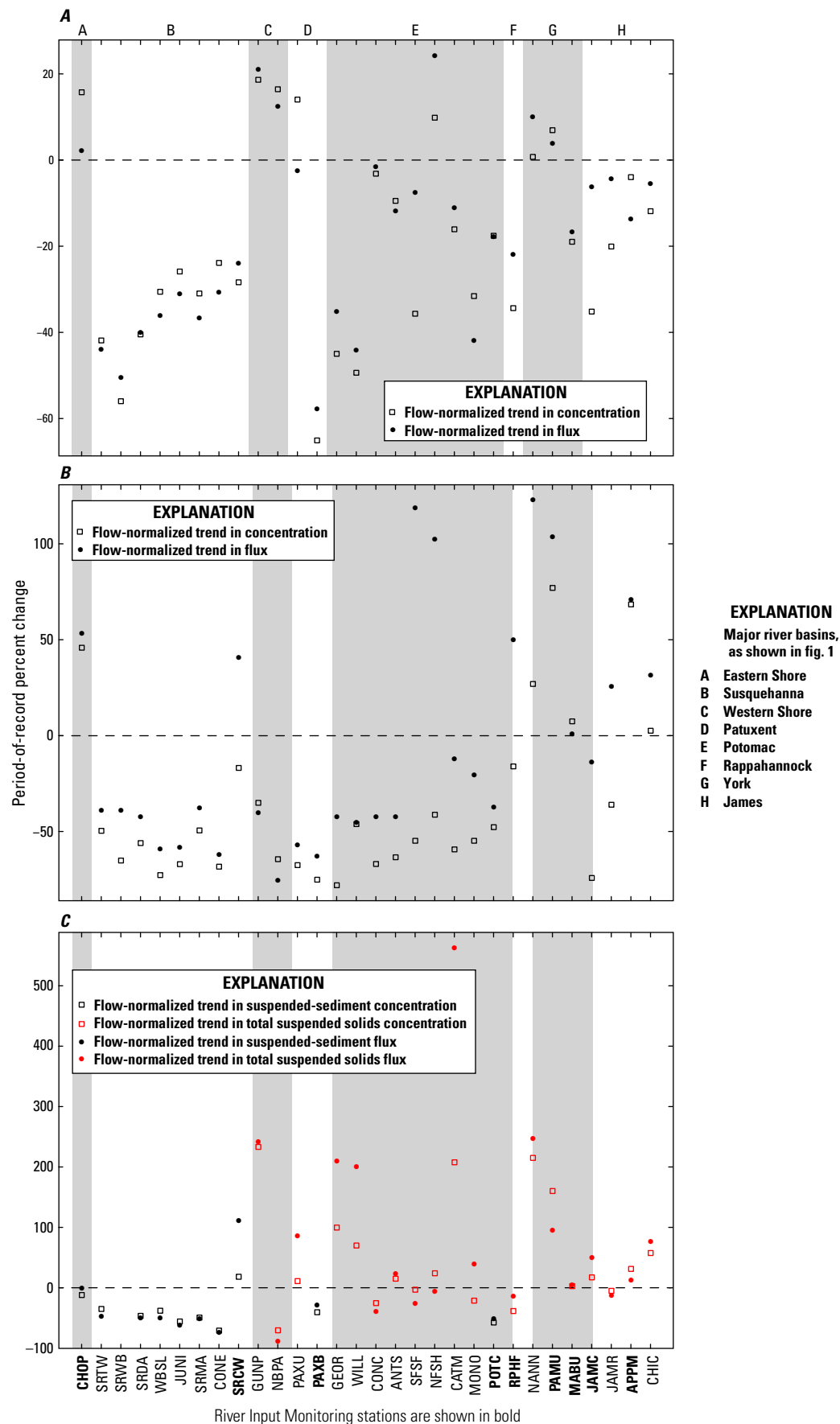
Trends

As discussed previously in the section “Comparison of ESTIMATOR and WRTDS Models on Chesapeake Bay Nontidal Water-Quality Monitoring Network Data Through Water Year 2012,” the capacity of the WRTDS model to independently estimate trends in both concentration and flux broadens the scope for interpretation of watershed-scale nutrient and sediment data. In this section of the report, those differences are summarized graphically for both *LONG_TERM* and *10_YEAR* (including *10_YEAR_SECONDARY*) scenarios, in terms of a percentage change relative to the simulation starting year. Tabular results for all graphics presented in this section, including period-of-record and 10-year trends for concentration, flux, and yield, expressed both in percentage and concentration/flux/yield units, are included in appendix 2 of this report. The quantitative interpretation of both the graphical and tabular data is hampered to some degree by the absence of confidence bounds on trends; thus, the following discussion focuses somewhat more on broad patterns and on the most evident contrasts between trends in flow-normalized concentration and trends in flow-normalized flux.

Period-of-Record Trends

Period-of-record trends in flow-normalized total nitrogen concentration, illustrated graphically in figure 18A and tabulated in appendix 2, were generally downward across the CBNTN. The minimum, median, and maximum trends in concentration were -65.1, -22.0, and +18.6 percent, respectively; only 7 out of 30 point estimates of long-term trends showed an increase (fig. 18A). Among the nine RIM stations, only the Choptank River near Greensboro, MD (CHOP, USGS station 01491000), and the Pamunkey River near Hanover, VA (PAMU, USGS station 01673000), showed upward trends in flow-normalized concentration, with 15.7 and 6.9 percent increases, respectively. The largest period-of-record decrease in concentration among the RIM stations occurred at the Patuxent River near Bowie, MD (PAXB, USGS station 01594440), with the 2012 concentration representing a 65.1 percent decrease relative to 1985. Period-of-record trends in flux coarsely tracked trends in concentration, with minimum, median, and maximum values of -57.7, -15.2, and +24.3 percent, respectively. Only one station, the Patuxent River near Unity, MD (PAXU, USGS station 01591000), had concentration and flux trends that differed in sign, showing an increasing concentration trend, but a decreasing (or more likely, insignificant) flux trend (fig. 18A).

The overall range of trends in flow-normalized total phosphorus concentration, with minimum, median, and maximum values of -78.1, -52.3, and +77.0 percent, respectively (fig. 18B), was generally larger than the range observed for total nitrogen. In addition, larger discrepancies between concentration and flux trends were evident for total phosphorus: minimum, median, and maximum flux trend values were -75.5, -37.6, and +122.9 percent, respectively. Point estimates of trend in flow-normalized concentration were positive at only 6 of the 30 long-term stations for which estimates were made. In contrast, estimated period-of-record trends in flux were positive at 11 of the 30 stations. The contrast between concentration and flux trends was most pronounced for the South Fork Shenandoah River at Front Royal, VA (SFSF, USGS station 01631000), and the North Fork Shenandoah River near Strasburg, VA (NFSH, USGS station 01634000), where point estimates of trend in flow-normalized concentration were negative, but corresponding estimates of trend in flow-normalized flux were about +100 percent (fig. 18B). As discussed by Moyer and others (2012), such a contrast is often indicative of downward trends in concentration throughout most of the discharge range, with an upward trend in concentration at the highest discharges. Among the RIM stations, similar contrasts were observed at the Susquehanna River at Conowingo, MD (SRCW, USGS station 01578310), and the Rappahannock River near Fredericksburg, VA (RPHF, USGS station 01668000); the James River at Cartersville, VA (JAMC, USGS station 02035000),



also showed a large difference in trend magnitude, although both trends were negative. As was the case for total nitrogen, the Patuxent River near Bowie, MD (PAXB, USGS station 01594440), had the most pronounced negative trends, in both concentration (−75.2 percent) and flux (−62.8 percent), of all the RIM stations.

Period-of-record flow-normalized trends in sediment concentration, representing a mixture of results for both total suspended solids and suspended sediment, showed the greatest range of the three constituents discussed in this section, with minimum, median, and maximum concentration trends of −71.4, −3.7, and +232.6 percent, respectively, and minimum, median, and maximum flux trends of −89.3, −0.4, and +562.6 percent, respectively (fig. 18C). The results indicate generally decreasing flow-normalized trends in the Susquehanna River Basin, with the exception of Susquehanna River at Conowingo, MD (SRCW, USGS station 01578310); elsewhere the number of downward and upward trending results were approximately equal. A particularly large contrast between concentration and flux trends is evident for Catoctin Creek near Middletown, MD (CATM, USGS station 01637500), with an upward concentration trend of 207 percent and an upward flux trend of 562 percent. For the RIM stations, the flow-normalized concentration trend was most strongly upward at the Pamunkey River near Hanover, VA (PAMU, USGS station 01673000), with an upward trend of 159.6 percent. Somewhat uncharacteristically, the period-of-record flow-normalized flux trend at that station, 95.2 percent upward, was less positive than the concentration trend, suggesting that the upward trend in concentration is more pronounced at frequently occurring moderate flows, rather than at the highest flows. The most positive flux trend among the RIM stations was at the Susquehanna River at Conowingo, MD (SRCW, USGS station 01578310; +110.9 percent); the most negative trends in both flow-normalized concentration and flux among the RIM stations were at the Potomac River at Chain Bridge, at Washington, DC (POTC, USGS station 01646580), with a trend in concentration of −58 percent and a trend in flux of −51.2 percent.

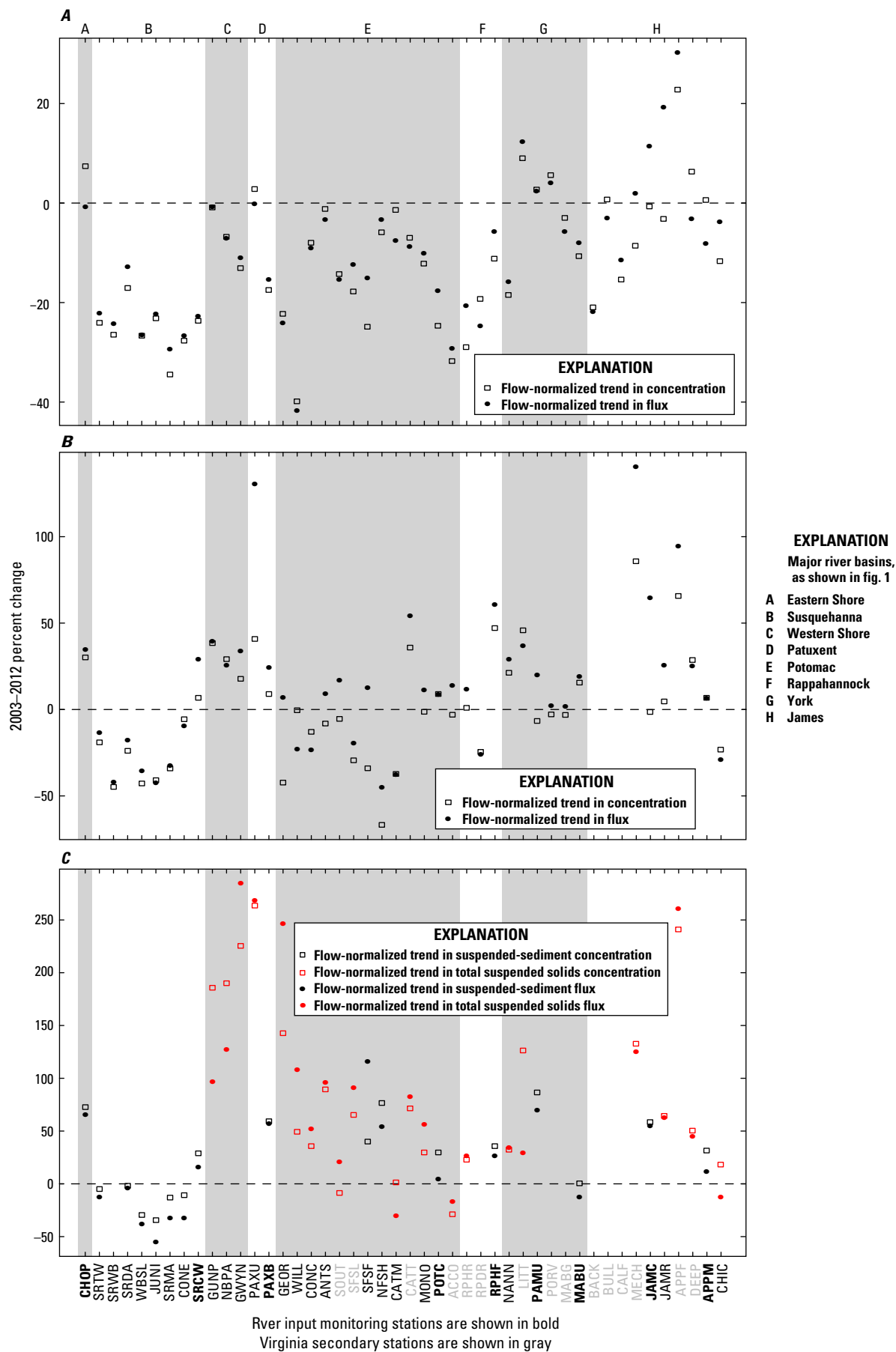
Ten-Year Trends

As was the case for period-of-record trends, the range of magnitudes for 10-year upward and downward flow-normalized trends was smallest for total nitrogen and largest for sediment. Overall, 10-year flow-normalized trends in total nitrogen were predominantly downward (fig. 19A), consistent with the pattern evident for period-of-record trends. The minimum, median, and maximum observed flow-normalized trends in concentration were −40.0, −12.7, and +22 percent, respectively; the median and range of flux trends were very

similar. Note that the increased number of results for 10-year trends in figure 19 is primarily due to the inclusion of the stations run under the *10_YEAR_SECONDARY* scenario (shown in the figure with gray station abbreviations). At the Choptank River near Greensboro, MD (CHOP, USGS station 01491000), the trend in concentration (+7.3 percent) was more positive than the trend in flux (−0.9 percent, indicating practically no trend); in contrast, at the James River at Cartersville, VA (JAMC, USGS station 02035000), the trend in concentration was slightly negative (or insignificant), while the trend in flux was +11.4 percent. Also among the RIM stations, the most pronounced negative trends in both flow-normalized concentration and flux were observed at the two largest tributaries: the Susquehanna River at Conowingo, MD (SRCW, USGS station 01578310), with concentration and flux trends of −24.2 and −23.3 percent, respectively, and the Potomac River at Chain Bridge, at Washington, DC (POTC, USGS station 01646580), with concentration and flux trends of −24.8 and −17.7 percent, respectively.

Ten-year flow-normalized trends in total phosphorus concentration ranged from −66.8 percent to +85.9 percent, with a median of −1.40 percent. Among the RIM stations, trends were most positive at the Rappahannock River near Fredericksburg, VA (RPHF, USGS station 01668000; +47.2 percent; fig. 19B), and the Choptank River near Greensboro, MD (CHOP, USGS station 01491000; +30.2 percent). Flow-normalized trends in flux had minimum, median, and maximum values of −44.9, +11.9, and +140.6 percent, respectively. Among RIM stations, trends were most positive for the James River at Cartersville, VA (JAMC, USGS station 02035000; +64.8 percent), and the Rappahannock River near Fredericksburg, VA (RPHF, USGS station 01668000; +61.0 percent). None of the RIM stations showed an appreciable downward 10-year trend in concentration or flux of total phosphorus.

Ten-year trends in sediment were generally upward, with downward trends in concentration at only 8 of the 39 stations for which trends were reported, and downward trends in flux at 10 stations (fig. 19C). The minimum, median, and maximum concentration trends were −34.4, +40.1, and +263.9 percent; corresponding values for trend in flux were −55.0, +52.2, and +285 percent. The Choptank River near Greensboro, MD (CHOP, USGS station 01491000), the Patuxent River near Bowie, MD (PAXB, USGS station 01594440), the Pamunkey River near Hanover, VA (PAMU, USGS station 01673000), and the James River at Cartersville, VA (JAMC, USGS station 02035000), had the most positive sediment trends among the RIM stations, with concentration trends between about 55 and 90 percent and flux trends between about 55 and 70 percent. The Mattaponi River near Beulahville, VA (MABU, USGS station 01674500), was the only RIM station with a decreasing sediment trend (−12.5 percent for flow-normalized flux; fig. 19C).



Selected Technical Considerations Associated With Operational Deployment of WRTDS Across the Chesapeake Bay Nontidal Water-Quality Monitoring Network

Inconsistencies in sampling and laboratory analysis protocols are typical features of many large, long-term water-quality networks. In the remainder of this report, the effects of these inconsistencies on flux and trend results produced using WRTDS are investigated, at least to a preliminary degree, using CBNTN data. Records with dense sub-daily sampling, variability in annual sampling effort, and variability in storm sampling effort are considered, and the effects of some common instances of changes in laboratory methods are explored.

Dense Sub-Daily Sampling

In choosing a daily time scale for WRTDS (and ESTIMATOR), it is assumed that concentration does not fluctuate significantly over the course of a day; that is, a single concentration measurement is representative of the entire day. This assumption tends to be more appropriate for very large rivers and less appropriate, if not clearly violated, in smaller streams. Because data for a given watershed can be provided

from multiple agencies, or may have been historically gathered at high frequency for specific studies, some CBNTN datasets have instances of two or more samples having been collected in a single day (fig. 20). As an extreme case, 46 percent of all samples collected at the Conestoga River at Conestoga, PA (USGS station 01576754), represent instances of sub-daily sampling (fig. 21). WRTDS and ESTIMATOR model developers considered four options for handling instances of sub-daily sampling: (1) retaining all samples, (2) retaining the highest- and lowest-concentration sample on any day sampled more than twice, (3) retaining a single interval-censored value representing the range of observed concentrations on that day, and (4) retaining a single observation representing the median of the observed values. The developers of both models recommended that, among these four alternatives, retaining a single median value provided the best representation of the true uncertainty in concentration on days with more than one sample (Robert Hirsch, U.S. Geological Survey, written commun., April 24, 2014). WRTDS results through water year 2012 had already been computed using option 2 above and published online prior to the issue of this recommendation. All results tabulated in the section “WRTDS Estimated Flux and Trend Results for Nutrients and Sediment for the Chesapeake Bay Nontidal Water-Quality Monitoring Network Through Water Year 2012” above reflect the use of option 2 above, consistent with published results. The recommended modification will be implemented in the production of all CBNTN products for the 2014 reporting cycle.

Figure 19. (facing page) Plots comparing point estimates of 10-year trends in flow-normalized concentration with corresponding trends in flow-normalized flux, in percent, for stations modeled under the *10_YEAR* scenario through 2012. *A*, Total nitrogen. *B*, Total phosphorus. *C*, Sediment. Shading patterns, identified with letters A–H, delineate major river basins, as shown in figure 1. Station abbreviations are identified in table 1.

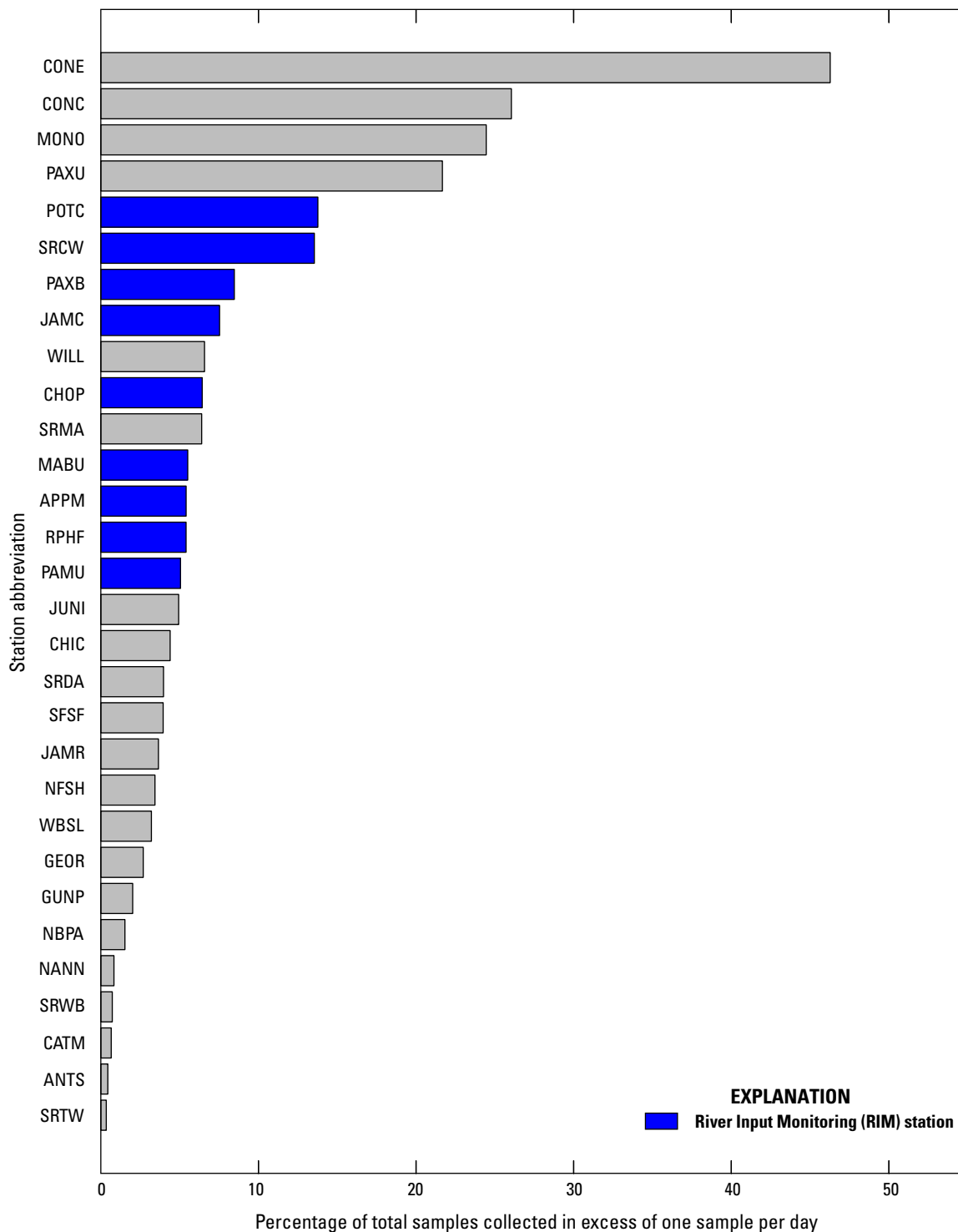


Figure 20. Chesapeake Bay Nontidal Water-Quality Monitoring Network (CBNTN) long-term stations, ranked in decreasing order of the percentage of total water-quality samples collected in excess of one sample per day. Station abbreviations are identified in table 1.

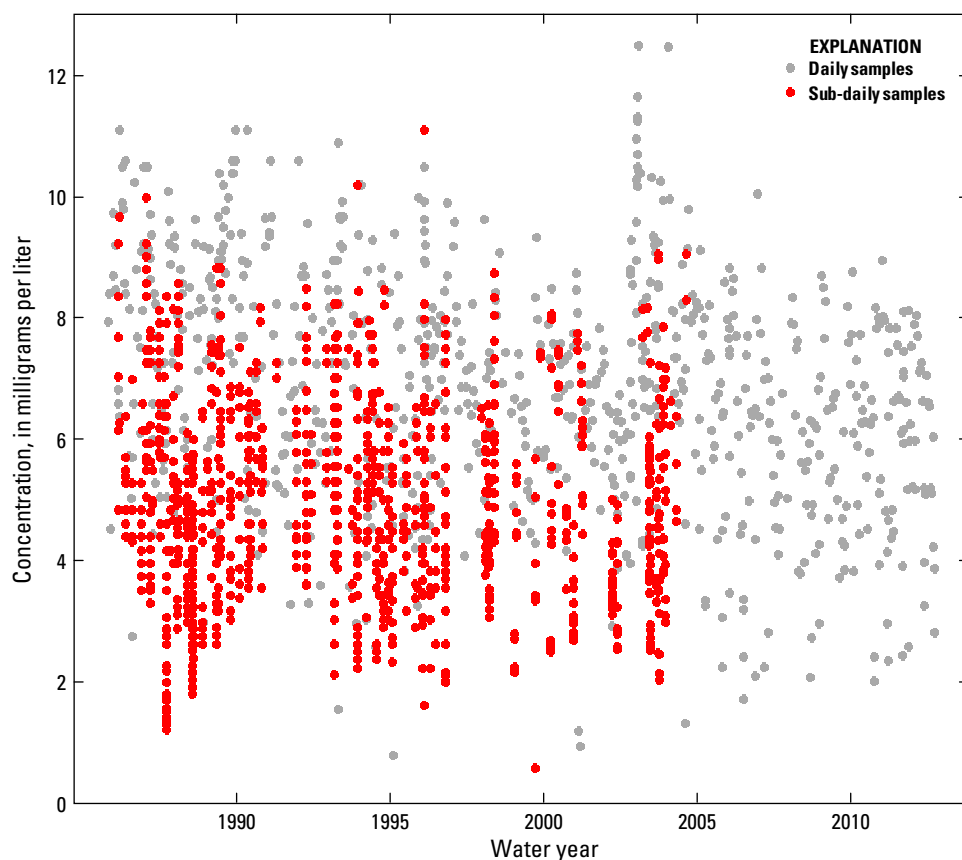


Figure 21. Plot of dissolved inorganic nitrogen in relation to date for the Conestoga River at Conestoga, PA, highlighting instances of sub-daily sampling.

Variability in Annual Sampling Effort and Storm Sampling Effort

As indicated in figure 20, stations where dense sub-daily sampling is a potential concern represent a rather small percentage of the long-term stations across the network: of 30 stations, only 4 have more than 15 percent of their total samples collected at a frequency greater than 1 per day. Several other sampling-related features of the records are more widespread, but the potential of those features to induce artifacts in trends have not been documented. Two features that will be examined in this section are year-to-year variability in overall sampling effort (that is, the total number of samples collected in each year) and year-to-year variability in storm sampling effort (that is, the total number of samples collected above some threshold constituting “high flow”). Figure 22 shows the period-of-record average number of samples per year collected at each CBNTN station having records exceeding 20 years. Design guidelines for the network prescribe 12 monthly samples and 8 targeted storm samples

per year for a total of 20 samples per year; actual averages shown in figure 22 range from under 10 samples per year for the Susquehanna River at Wilkes-Barre, PA (SRWB, USGS station 01536500), to nearly 40 samples per year for the Potomac River at Chain Bridge, at Washington, DC (POTC, USGS station 01646580). High average annual sample counts may be the result of multiple agencies collecting similar data or samples collected in support of targeted process studies. Figure 23 shows annual sample counts for 8 of the 12 stations having the highest annual averages. Periods of dense sampling may persist for years and appear/disappear abruptly, or they may last for only one or a few years. In contrast, plots showing annual sample counts over time for the eight stations having the lowest average annual sample counts are shown in figure 24. This figure indicates that stations having the lowest average annual sample counts share a common cause: these are stations that transitioned from strictly monthly samples for most of their period-of-record to the prescribed monthly-plus-storm protocol upon becoming a part of the CBNTN. The figure indicates that most of the stations that made this transition did so around 2006.

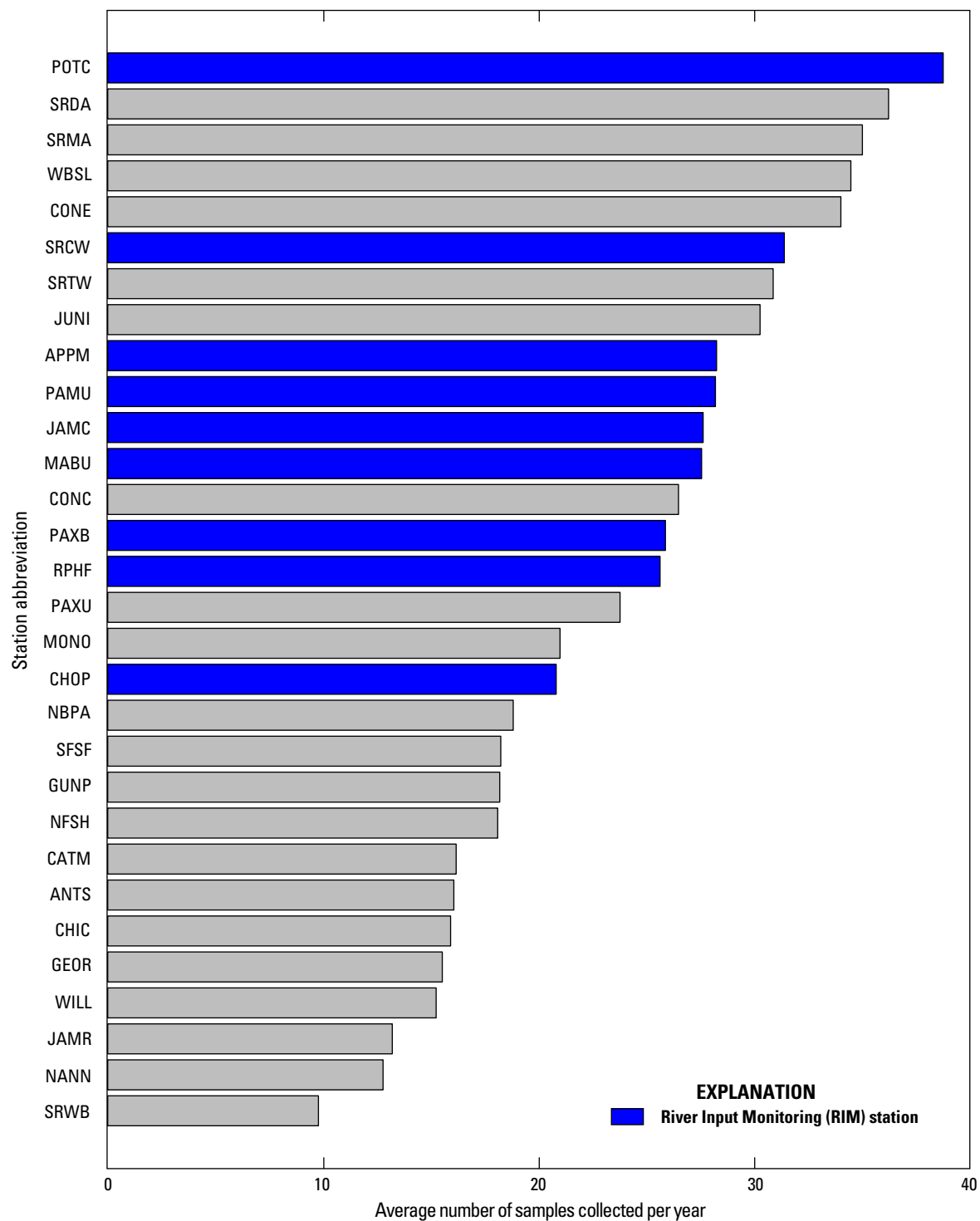


Figure 22. Chesapeake Bay Nontidal Water-Quality Monitoring Network (CBNTN) long-term monitoring stations, ranked in decreasing order of average annual sampling frequency. Station abbreviations are identified in table 1.

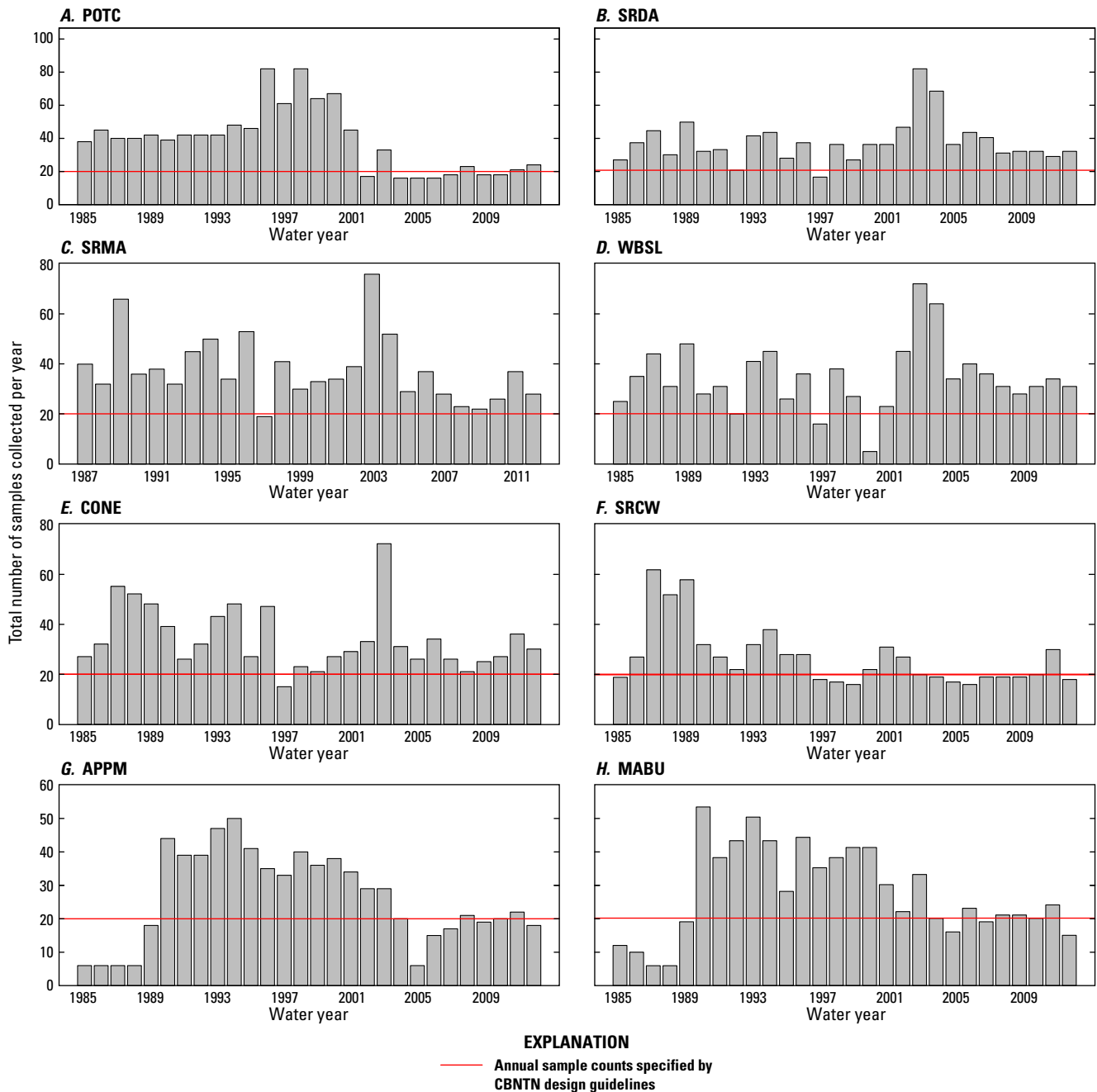


Figure 23. Number of samples collected per year at 8 of the 10 Chesapeake Bay Nontidal Water-Quality Monitoring Network (CBNTN) long-term stations with the highest average annual sampling frequency. Station abbreviations are identified in table 1.

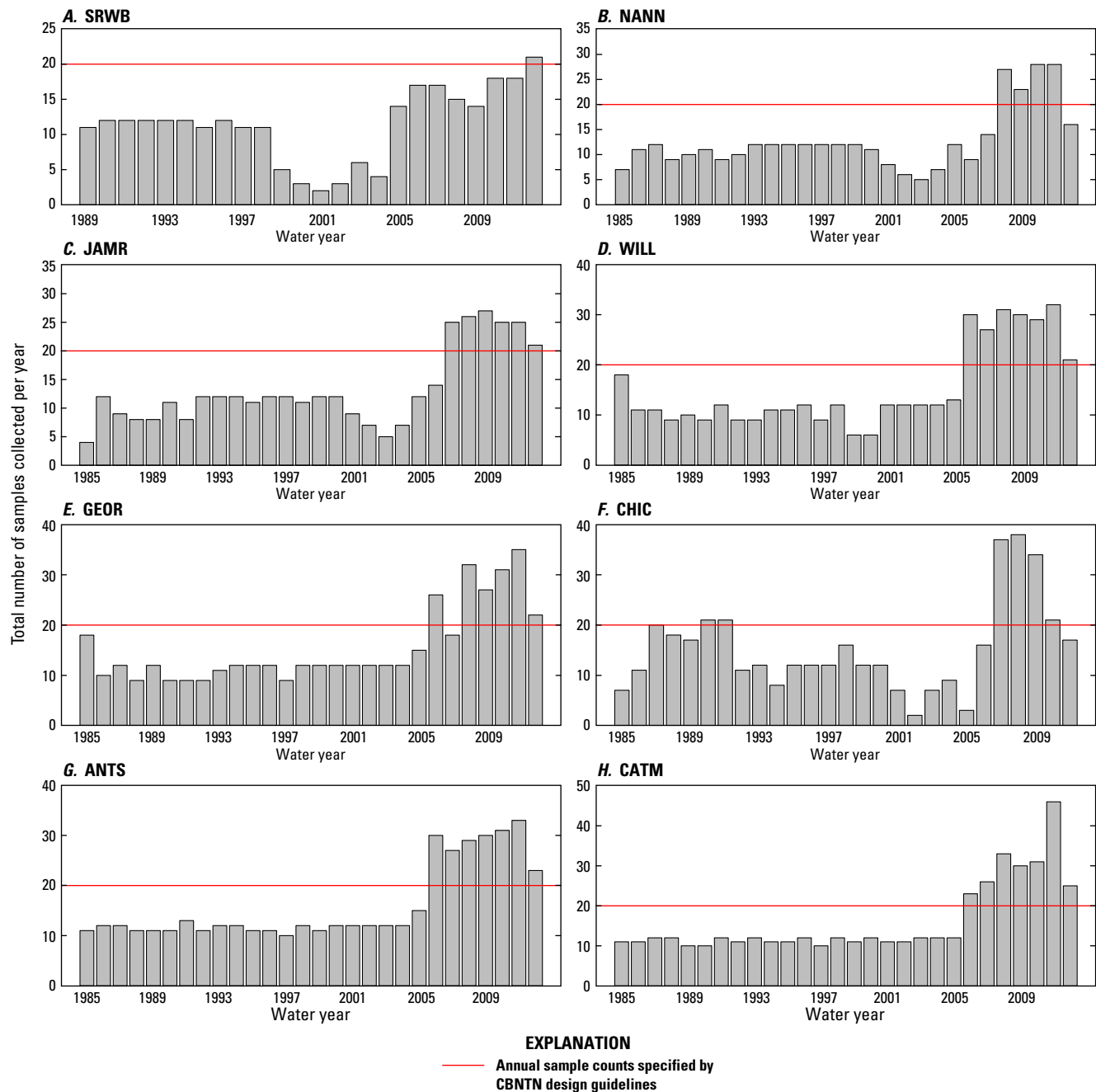


Figure 24. Number of samples collected per year at the eight Chesapeake Bay Nontidal Water-Quality Monitoring Network (CBNTN) long-term stations with the lowest average annual sampling frequency. Station abbreviations are identified in table 1.

Sprague (2001) conducted a study of the effects of storm sampling frequency on flux and trend estimation by applying ESTIMATOR to manipulated 5- and 10-year water-quality records for the Rappahannock River near Fredericksburg, VA (RPHF, USGS station 01668000), and the James River at Cartersville, VA (JAMC, USGS station 02035000). Samples collected at these stations were classified as either “stormflow” or “base flow” depending on the percentage of base flow, as determined by a streamflow-separation algorithm, in total measured streamflow; datasets were then generated incorporating all base-flow samples but varying densities of storm samples. Fluxes and trends estimated using these subsamples were then compared to “baseline” fluxes and trends estimated using the complete dataset. For flux estimates, Sprague (2001) found that for dissolved constituents, fewer storm samples were required to achieve a given degree of match between estimates based on subsamples and the baseline than was the case for particulate constituents. Trend estimates, which were based only on the linear time term in equation 1, were found to be generally less sensitive to storm sampling frequency than flux estimates, but the short records and absence of strong trends in the baseline data complicated determination of the relative sensitivity of dissolved and particulate constituents.

In light of the documented importance of characterizing the concentration-discharge relation over the widest possible range of daily discharges, it is reasonable to question whether trends reported at stations such as those shown in figure 24 represent true environmental conditions, or whether they are artifacts induced by the introduction of storm sampling in about 2006. At the most basic level, the prescribed targeting of high flows, essential for completely characterizing the region of the concentration-discharge relation most important for predicting high fluxes, is missing for more than two-thirds of the record. It can be argued, at least in principle, that if those high-flow concentrations are unknown, resulting estimated fluxes and, by extension, trends in fluxes that reach into the period of prescribed targeted sampling are suspect. A somewhat drastic resolution in such a case would be to truncate the record and consider only data collected since the prescribed protocol was implemented. On the other hand, monthly sampling might in some case produce enough high-flow information that differences between a “minimally characterized” and “thoroughly characterized” concentration-discharge relation are small relative to sampling variability. If that is the case, then discarding the bulk of the record, which may have valuable long-term information at all but the most extreme flows, would certainly be undesirable. To better understand the potential influence of variability in storm sampling effort on estimated trends using WRTDS, a subsampling experiment was conducted using eight of the most richly sampled stations in figure 23 as “baseline” datasets. In order to more effectively isolate the effects of changes in storm sampling protocol from variability that could be expected from down-sampling a dense dataset, this experiment was designed to control for variability

in overall annual sample count, as well as for variability in storm sampling frequency. The sections that follow document the effect of these two heterogeneities separately. In general, “thinning” a record with periods of dense annual sampling down to a consistent 20 samples per year, if done in a manner that generally preserves the baseline frequency of high-flow sampling, is shown to have a rather small influence on WRTDS estimates of trend in concentration or flux. Further manipulation of the thinned records to mimic the initiation of storm sampling in a long record of monthly-only samples indicates that changes in high-flow sampling protocol can result in more appreciable distortion of estimated trends in flow-normalized flux, especially for phosphorus and sediment.

Description of Methods

A set of densely sampled stations was repeatedly subsampled to produce 10 water-quality datasets consisting of the design guideline 20 samples per year, preserving as closely as possible the year-to-year variability in “high-flow” sampling frequency present in the original data. This process was accomplished in a manner that identified each sample in the subsampled data as a “monthly” or “storm” sample. The effects of variability in overall annual sampling frequency on reported trends were quantified using the design guideline (for example, “monthly” plus “storm”) subsamples. WRTDS was run on each of those subsamples, and the variability of estimated trends between subsamples, and in relation to the trend based on the “baseline” dataset consisting of all observations, was quantified. The effects of variability in storm sampling frequency on reported trends were quantified by further manipulating the 10 design guideline subsamples. For each subsample, “storm” samples were deleted from the record in all years prior to water year 2003, resulting in datasets with 12 samples per year through 2003, and 20 samples per year thereafter. Trends estimated using these heterogeneous records were compared to trends estimated using the design guideline records with storm samples retained throughout the record.

For the baseline datasets, eight stations were chosen from those having the highest overall sampling frequency. Seven of the eight most intensely sampled stations are in the Susquehanna River Basin; the exception was the Potomac River at Chain Bridge, at Washington, DC (POTC, USGS station 01646580; fig. 22). Because the Susquehanna River stations were all sampled by the same agencies and have similar patterns of time variability, the Mattaponi River near Beulahville, VA (MABU, USGS station 01674500), and the Appomattox River at Matoaca, VA (APPM, USGS station 02041650), were substituted for the Susquehanna River at Towanda, PA (SRTW, USGS station 01531500), and the Juniata River at Newport, PA (JUNI, USGS station 01567000), in order to better represent patterns in time variability across the network (fig. 23). The subsampling outlined above was then performed on each of the eight datasets as follows:

- Each year was divided into two 6-month seasons; all the samples collected in a given year/season formed the pool of “candidate” samples for the same 6-month period in each subsample.
- For each month in the 6-month season, a sample was drawn from the subset of “candidate” pool samples collected in that month. The sample was drawn at random and without replacement, but was subject to the qualification that the associated daily flow be at or below a specified percentile of the daily flows for the given month. (An exception to this rule was that if only one sample were available for the month, that sample was taken without qualification.) These samples were identified in the subsample as “Monthly.” The parameter governing the upper threshold percentile of daily flows from which a “Monthly” sample was drawn is referred to as the “upperMonthly” threshold in table 4—a single threshold applied to all 12 months over the entire period of record.
- Once samples were drawn in this manner for each month in the 6-month period, four storm samples were drawn from the samples remaining in the pool. This was also done at random and without replacement, but was subject to the restriction that the associated daily flow be at or above a specified percentile of the daily flows for each 6-month season. These samples were identified in the subsample as “Storm.” The parameter governing the lower threshold percentile of daily flows from which a “Storm” sample was drawn is referred to as the “lowerStorm” threshold in table 4—a single threshold applied to both 6-month seasons over the entire period of record.

The effect of the “upperMonthly” parameter was to exclude samples collected at the highest daily flows of any month from consideration as a “monthly” sample; correspondingly, the effect of the “lowerStorm” parameter was to exclude remaining samples collected at the lowest daily flows of any season from consideration as a “storm” sample. Because the eight baseline datasets represented different histories of data collection and station-specific study, the datasets varied widely both in terms of annual sample counts and the frequency with which high flows were sampled. In light of this variation, generating subsampled datasets that allowed for independent evaluation of the effects of “thinning” a densely sampled baseline down to the design guideline 20 samples per year and the effects of initiating targeted storm sampling to a record previously consisting only of monthly samples required that the two subsampling parameters be “tuned” for each watershed (table 4). This process was done according to two criteria:

1. Averaged over the period of record, the between-watershed pattern of the percentage of “monthly” plus “storm” samples collected at high flows coarsely resembled the between-watershed pattern in the baseline datasets; for this purpose, high flows are defined as the 95th percentile of daily flows observed for the given watershed in the given subsampled year.
2. When only “monthly” samples were considered, that is, when samples identified as “storm” samples were removed from the entire record, the period-of-record average frequency of samples collected at high flows, defined as above, in each watershed was about 5 percent, the coarse expectation if monthly sampling were truly flow independent.

Table 4. Subsampling algorithm parameters for each of the eight long-term stations used in the subsampling experiment.

Station abbreviation	Station ID	Station name	Subsampling parameter ¹	
			upperMonthly (percentile)	lowerStorm (percentile)
POTC	01646580	Potomac River at Chain Bridge, at Washington, DC	0.97	0.25
SRDA	01540500	Susquehanna River at Danville, PA	0.90	0.25
SRMA	01576000	Susquehanna River at Marietta, PA	0.80	0.25
WBSL	01553500	West Branch Susquehanna River at Lewisburg, PA	0.90	0.25
CONE	01576754	Conestoga River at Conestoga, PA	0.95	0.25
SRCW	01578310	Susquehanna River at Conowingo, MD	0.70	0.25
APPM	02041650	Appomattox River at Matoaca, VA	0.85	0.25
MABU	01674500	Mattaponi River near Beulahville, VA	0.80	0.25

¹Refer to section “Variability in Annual Sampling Effort and Storm Sampling Effort” for parameter definitions.

An illustration of the procedure for drawing a single subsample from the baseline data for total phosphorus at the Susquehanna River at Marietta, PA (SRMA, USGS station 10576000), is provided in figure 25. Figure 25A shows the total number of samples in each year that were collected at this station. Figure 25B shows the proportion of samples in each year that were collected at or above the 95th percentile of daily flows for that year; the red line in figure 25B reflects the period-of-record average, just over 20 percent. Figure 25C shows the sample counts for pooled monthly and storm samples for a single design guideline subsample after applying the procedure described above, using the upperMonthly and lowerStorm parameters tabulated for the Susquehanna River at Marietta in table 4. The shortfall below 20 samples in some years is due to two factors: First, some months were not sampled in the baseline and thus could not be represented in the subsample. Second, for some periods, once all monthly samples were drawn from the candidate pool for a given 6-month season, fewer than four samples above the specified lowerStorm threshold remained for consideration as candidate storm samples. As indicated in figure 25D, the long-term average proportion of samples collected at or above the 95th percentile of daily flows for each year was about 20 percent, similar to the frequency in the baseline data. Figure 25E indicates the annual sample count after storm samples are removed; shortfalls below 12 samples in this series are due solely to the fact that some months had no samples in the baseline data. This series of monthly samples is shown only for purposes of illustrating the subsampling method; the heterogeneous subsample corresponding to the design guideline subsample shown in figure 25C would have sample counts as shown in figure 25E prior to 2003 and sample counts as shown in figure 25C thereafter. Finally, figure 25F indicates that the long-term average proportion of monthly samples collected at or above the 95th percentile of daily flows for each year for this particular subsample was about 4 percent. The long-term average high-flow percentages shown in figures 25D and 25F were both achieved by hand-tuning the upperMonthly and lowerStorm parameters for the Susquehanna River at Marietta to a single pair of values; the same values were used in selecting all 10 subsamples for each watershed (table 4). Summary statistics for all 10 subsamples at all 8 watersheds (fig. 26) indicate that the design guideline subsamples including monthly plus storm samples broadly replicate the between-watershed pattern of average high-flow sampling frequency and that average high-flow sampling frequency for the same subsamples with storm samples deleted is approximately 5 percent.

In the interest of characterizing the effects of thinning and initiating storm sampling for both dissolved and particulate constituents, the subsampling experiment was repeated, across all eight stations, for dissolved inorganic nitrogen, total phosphorus, and sediment. Results for all three constituents are discussed in the following section.

Results: Variability in Annual Sampling Effort

At Susquehanna River at Marietta, PA (SRMA, USGS station 01576000), estimated trends for dissolved inorganic nitrogen using (a) the baseline data, (b) the annual median ($n=10$) of the subsampled trends, and (c) the range of the subsampled trends nearly coincide (fig. 27); visual inspection suggests that in any year the maximum deviation of concentration and flux between the subsampled annual flow-normalized estimates and the corresponding baseline estimates is well under 10 percent (in magnitude) of the estimate using the baseline data. The range in annual estimates for dissolved inorganic nitrogen concentration and flux based on the subsamples deviated from the trend estimated from the baseline data just after the middle of the record, coincident with the period of densest annual sampling in the baseline data. This is consistent with the notion that periods of densest sampling in the baseline data should correspond to periods of greater between-sample variability in the subsamples. For total phosphorus, the range of variability in flow-normalized concentration was greater than that observed for dissolved inorganic nitrogen (fig. 28A). Moreover, the range of variability in flow-normalized flux for total phosphorus was greater than that observed for concentration, particularly early in the record, where the maximum deviation between the subsampled flow-normalized estimates and the baseline estimate was about 20 percent of the estimate using the baseline data (fig. 28B). Recall that flow-normalized concentration tends to be driven by the concentration estimates throughout the range of discharges, whereas flow-normalized flux is sensitive primarily to the concentrations at high discharges. The concentration-discharge relation for total phosphorus at Susquehanna River at Marietta is parabolic upward, with highest concentrations at highest flows; in contrast, the relation for dissolved inorganic nitrogen is parabolic downward, with the highest concentration near the middle of the range of sampled discharges (data not shown). For total phosphorus (fig. 28), sampling variability for the high-flow observations, perhaps limited to the presence or absence of a few key observations in the subsample, is likely driving variability in estimated concentrations at the highest flows, thereby inducing a wider range of variability in flow-normalized flux, relative to flow-normalized concentration, than was observed for dissolved inorganic nitrogen (fig. 27). This phenomenon is evident to an even greater degree for suspended-sediment concentration (fig. 29)—generally the constituent with the strongest, most consistent upward-parabolic concentration-discharge relation of the three constituents considered in this analysis. For this constituent, the maximum deviation between the annual subsampled flow-normalized flux estimates and the corresponding baseline estimates was 35 percent of the estimate using the baseline data (fig. 29B).

Results associated with “thinning” baseline datasets for the other 7 stations down to 20 design guideline samples per year were broadly similar to those discussed above for

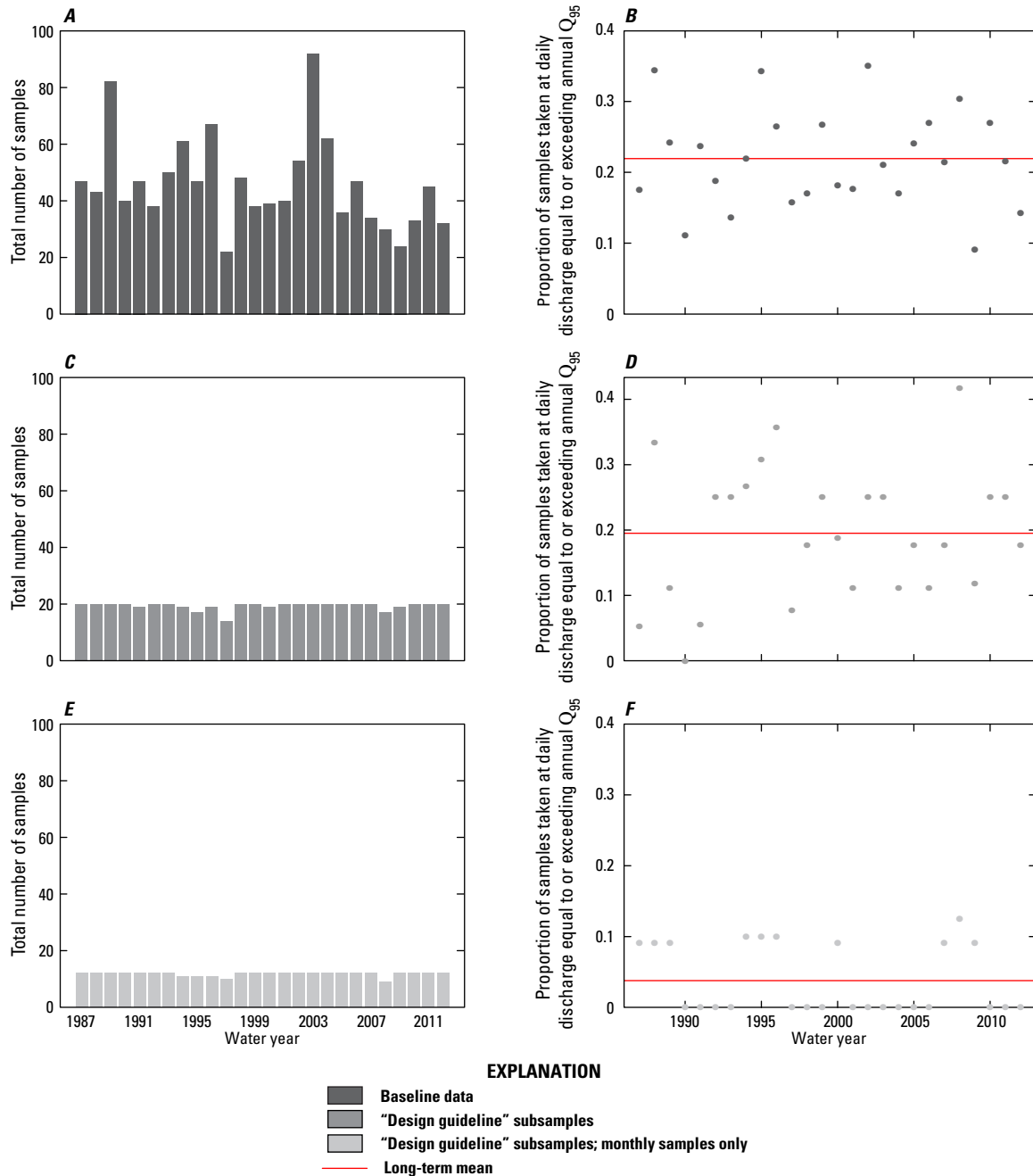


Figure 25. Comparison of overall annual sampling frequency and storm sampling frequency of a single subsample of the record for total phosphorus at the Susquehanna River at Marietta, PA, with corresponding values in the baseline dataset. *A*, Total number of samples per year in the baseline data. *B*, Proportion of baseline samples for each year collected at daily discharge at or above the year-specific daily Q_{95} . *C*, Total number of monthly plus storm samples in a single "design guideline" subsample. *D*, Proportion of monthly plus storm samples collected at daily discharge at or above Q_{95} in a single "design guideline" subsample. *E*, Total number of monthly samples in the "design guideline" subsample. *F*, Proportion of monthly samples collected at daily discharge at or above the Q_{95} in a single "design guideline" subsample. [Q_{95} is the 95th percentile of daily flows observed in a given year.]

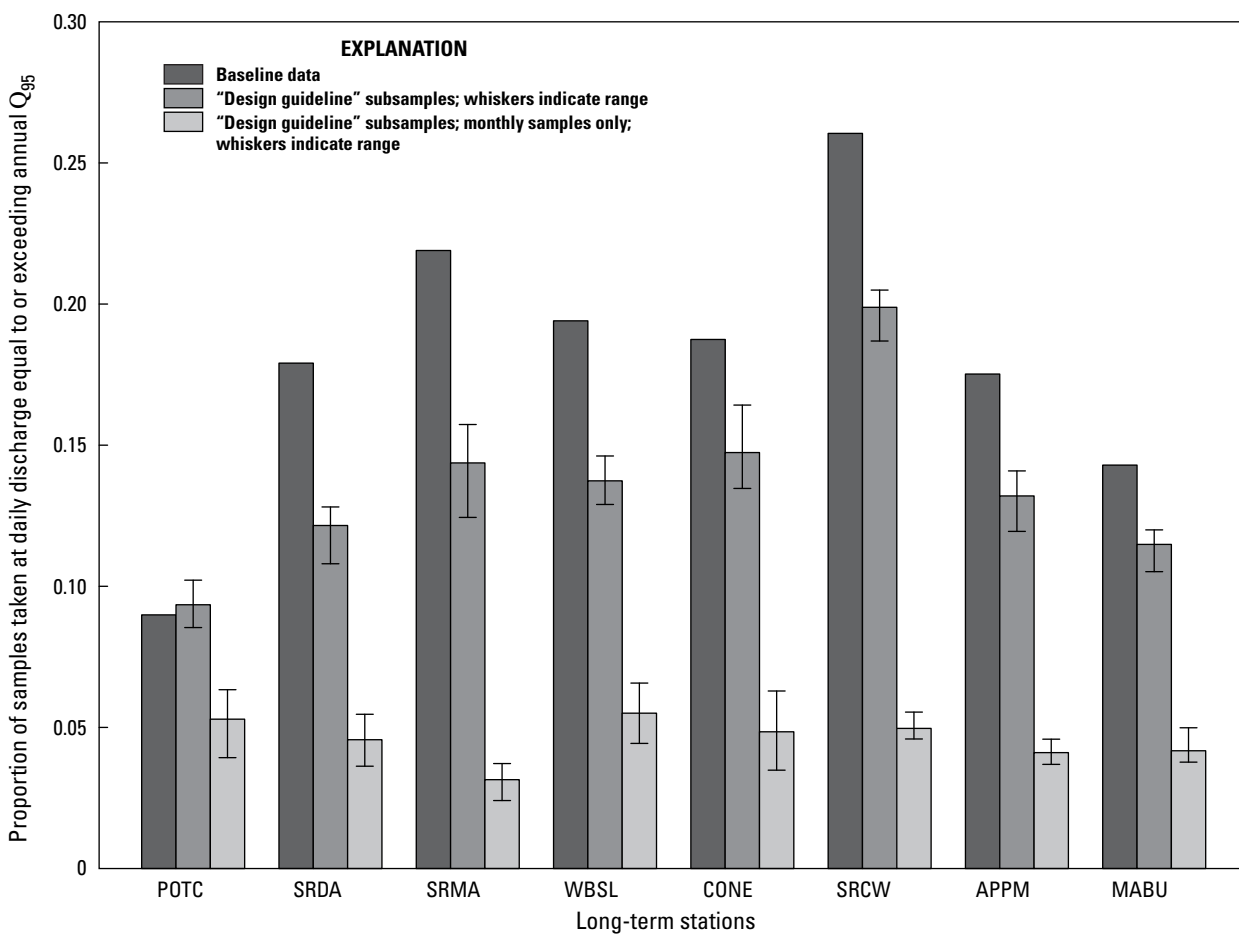


Figure 26. Storm sampling frequency summary statistics for all eight long-term stations used in the subsampling experiment. Station abbreviations are identified in table 1. [Q_{95} is the 95th percentile of daily flows observed in a given year.]

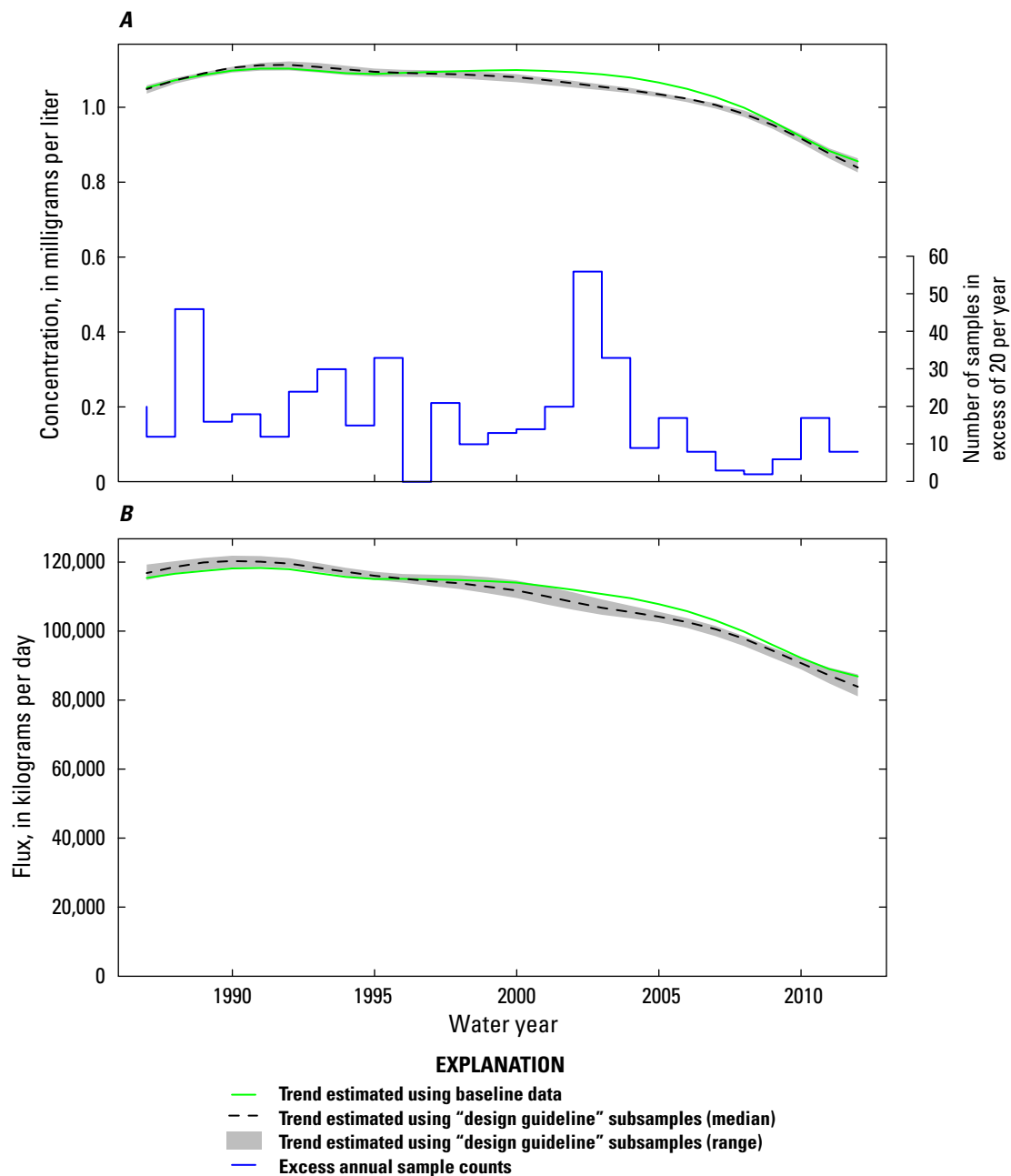


Figure 27. Long-term trends in flow-normalized concentration and flux of dissolved inorganic nitrogen at Susquehanna River at Marietta, PA, estimated using both the baseline record and 10 "design guideline" subsamples. *A*, Flow-normalized concentration. *B*, Flow-normalized flux.

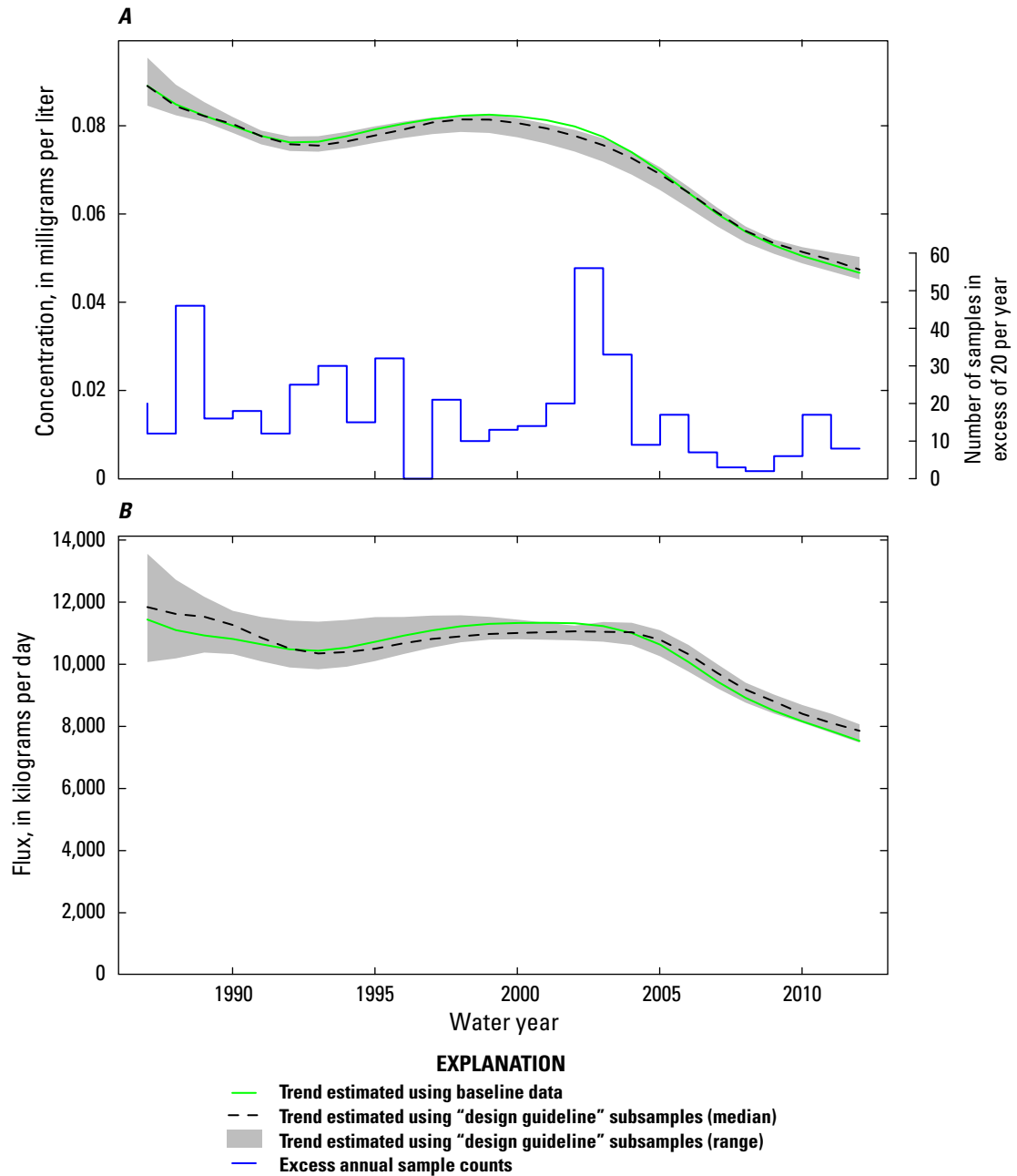


Figure 28. Long-term trends in flow-normalized concentration and flux of total phosphorus at Susquehanna River at Marietta, PA, estimated using both the baseline record and 10 "design guideline" subsamples. *A*, Flow-normalized concentration, with overlay showing relative baseline sampling density. *B*, Flow-normalized flux.

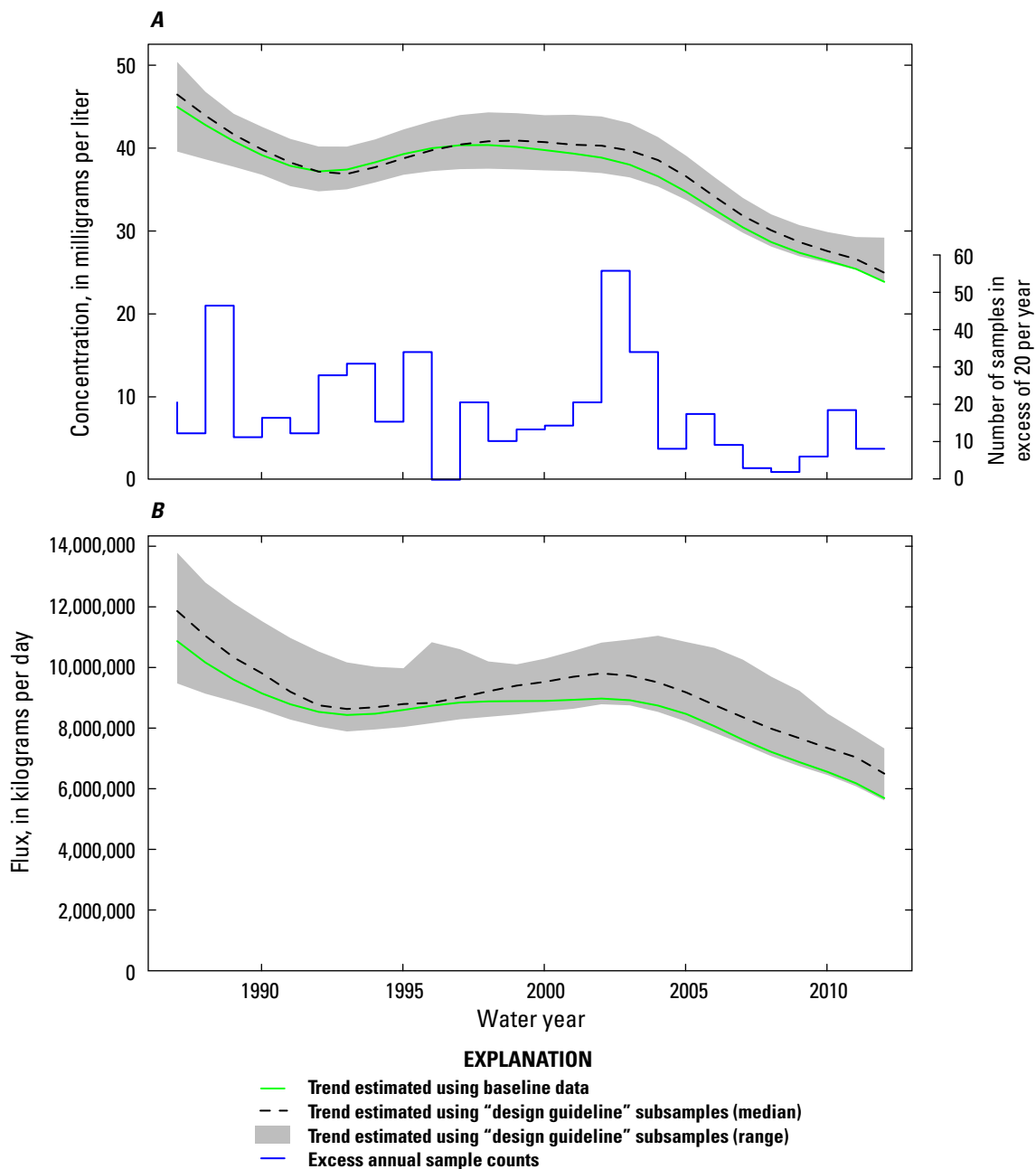


Figure 29. Long-term trends in flow-normalized concentration and flux of sediment at Susquehanna River at Marietta, PA, estimated using both the baseline record and 10 "design guideline" subsamples. *A*, Flow-normalized concentration, with overlay showing relative baseline sampling density. *B*, Flow-normalized flux.

the Susquehanna River at Marietta. Graphs for the other 7 stations, corresponding to figures 27–29 for the Susquehanna River at Marietta, are compiled in appendix 4. For dissolved inorganic nitrogen, the 8-station median of the maximum single-year relative deviation between flow-normalized flux estimates computed using design guideline subsamples and estimates for the corresponding year computed using the baseline data was 6 percent. Maximum relative differences for flow-normalized flux of total phosphorus and suspended sediment were more pronounced, with 8-station median values of 22 percent for phosphorus and 32 percent for sediment. The worst-case maximum single-year relative deviations between flow-normalized flux estimates computed using design guideline subsamples and estimates for the corresponding year computed using the baseline data were 16, 42, and 70 percent for dissolved inorganic nitrogen, total phosphorus, and sediment, respectively. Statistical hypothesis testing was not possible given the lack of uncertainty bounds on long-term trends; however, despite the observed deviations, there was little subjective evidence to suggest that a trend using a design guideline subsample from a densely sampled dataset would lead a land manager or researcher to a dramatically different conclusion about trend direction or shape than would have been reached using the baseline data.

Results: Variability in Storm Sampling Effort

For dissolved inorganic nitrogen at the Susquehanna River at Marietta, PA (SRMA, USGS station 01576000), the most noticeable effect of removing storm samples, prior to 2003, from the design guideline subsamples was to expand the range of variability in flow-normalized concentration and flux estimates, relative to the range of estimates obtained using the design guideline subsamples, in the early years of the record (fig. 30). (Note that in figure 30 and the two figures that follow, the trend line based on the baseline dataset is not shown.) Nevertheless, the maximum relative differences between flow-normalized estimates of both concentration and flux computed using heterogeneous subsamples and estimates for the corresponding year computed using the baseline data remained small (5 and 11 percent for concentration and flux, respectively), and the overall shape of the trends estimated using the heterogeneous subsamples closely tracked the shape originally estimated using the baseline data (fig. 27). The ranges of between-subsample variability in flow-normalized estimates of total phosphorus concentration and flux using the heterogeneous subsamples were comparable to the ranges observed using the design guideline subsamples. However, the range of values estimated for total phosphorus using the heterogeneous subsamples tended to be offset downward from the range estimated using the design guideline subsamples, to the extent that the ranges for trends in concentration completely diverged near the middle of the record (fig. 31A). Flow-normalized trends for sediment at the Susquehanna River at Marietta (fig. 32) displayed the greatest degree of sensitivity to heterogeneities in sampling protocol, with

maximum relative differences between estimates computed using heterogeneous subsamples and estimates for the corresponding year computed using the baseline data on the order of 25 percent for concentration and 45 percent for flux. For all three constituents examined in this experiment, the ranges in flow-normalized trends in both concentration and flux estimated using heterogeneous subsamples converged to the corresponding ranges estimated using the design guideline subsamples soon after water year 2003, the first year that each pair of corresponding subsamples contained the identical set of subsampled water-quality observations.

Graphs for the other seven stations, corresponding to graphs for the Susquehanna River at Marietta, PA (figs. 31–32), are compiled in appendix 5. For dissolved inorganic nitrogen, the 8-station median of the maximum single-year relative deviation between flow-normalized flux estimates computed using heterogeneous subsamples and estimates for the corresponding year computed using the baseline data was 15 percent. Maximum relative differences for flow-normalized flux of total phosphorus and suspended sediment were more pronounced, with 8-station median values of 30 percent for phosphorus and 53 percent for sediment. The worst-case maximum single-year relative deviations between flow-normalized flux estimates computed using heterogeneous subsamples and estimates for the corresponding year computed using the baseline data were 25, 37, and 250 percent for dissolved inorganic nitrogen, total phosphorus, and sediment, respectively.

The two primary uses for water-quality flux and trend estimates in the CBNTN are (1) assessment of progress in meeting regulatory nutrient and sediment loading reductions, and (2) scientific interpretation to better understand watershed-scale reaction/transport processes. The range of between-subsample variability evident for the heterogeneous subsamples in figures 30–32 and appendix 5, particularly for flow-normalized flux trends for the predominantly particulate constituents total phosphorus and sediment, indicates potential for distortion sufficient to mislead land managers and regulators regarding the magnitude of period-of-record change, and (or) set researchers on a search for process-based explanations for results that are in fact artifacts resulting from a change in sampling protocol. Given the tendency for trends estimated using heterogeneous subsamples to converge toward the trend estimated using the baseline data after 2003, one obvious option for minimizing the potential for appreciable artifacts when estimating flux and trends for particulate constituents is to compute the trends using the entire record but report them only for the period that design guideline sampling protocols were consistently followed. Another commonly proposed alternative to handling records with known heterogeneities is to truncate the record and analyze only the “consistent” portion. To test the effect of discarding data lacking storm samples, 10-year trends were estimated using only post-2003 data, the period over which the design guideline subsamples and heterogeneous subsamples contained exactly the same observations; these subsamples are referred to hereafter

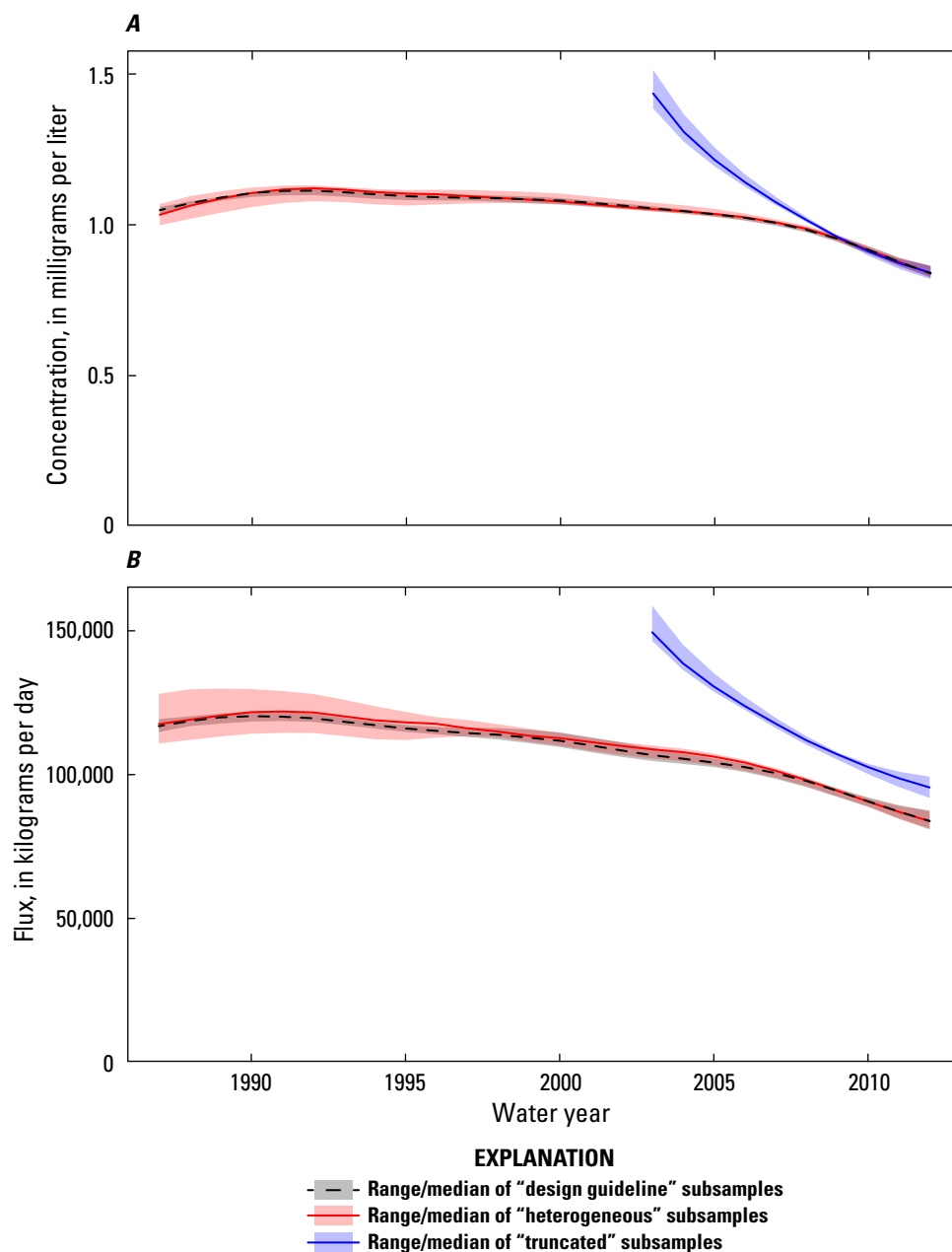


Figure 30. Long-term trends in flow-normalized concentration and flux of dissolved inorganic nitrogen at Susquehanna River at Marietta, PA, estimated using 10 “design guideline” subsamples, 10 “heterogeneous” subsamples, and 10 subsamples with the simulation period truncated to 2003–12. *A*, Flow-normalized concentration. *B*, Flow-normalized flux.

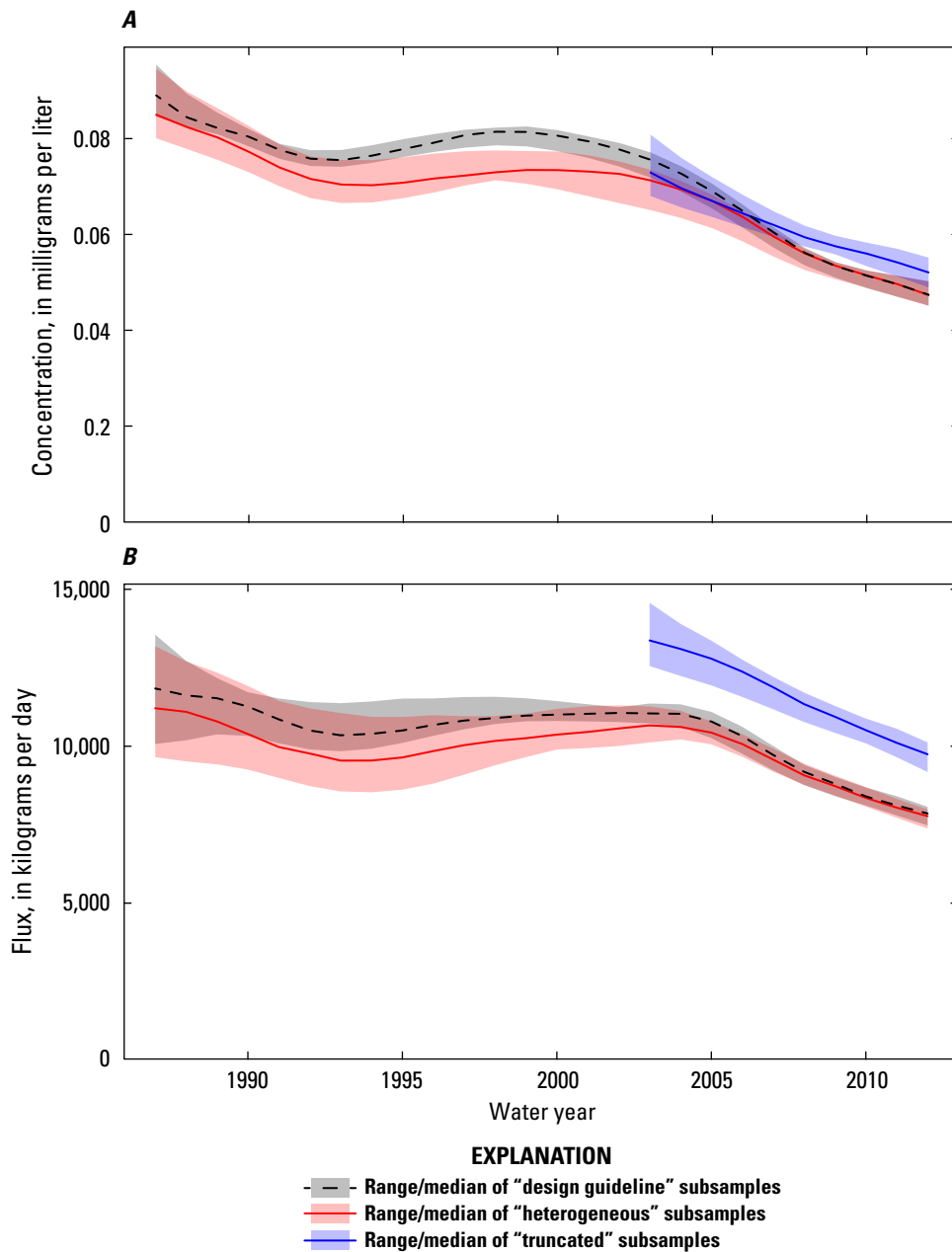


Figure 31. Long-term trends in flow-normalized concentration and flux of total phosphorus at Susquehanna River at Marietta, PA, estimated using 10 "design guideline" subsamples, 10 "heterogeneous" subsamples, and 10 subsamples with the simulation period truncated to 2003–12. *A*, Flow-normalized concentration. *B*, Flow-normalized flux.

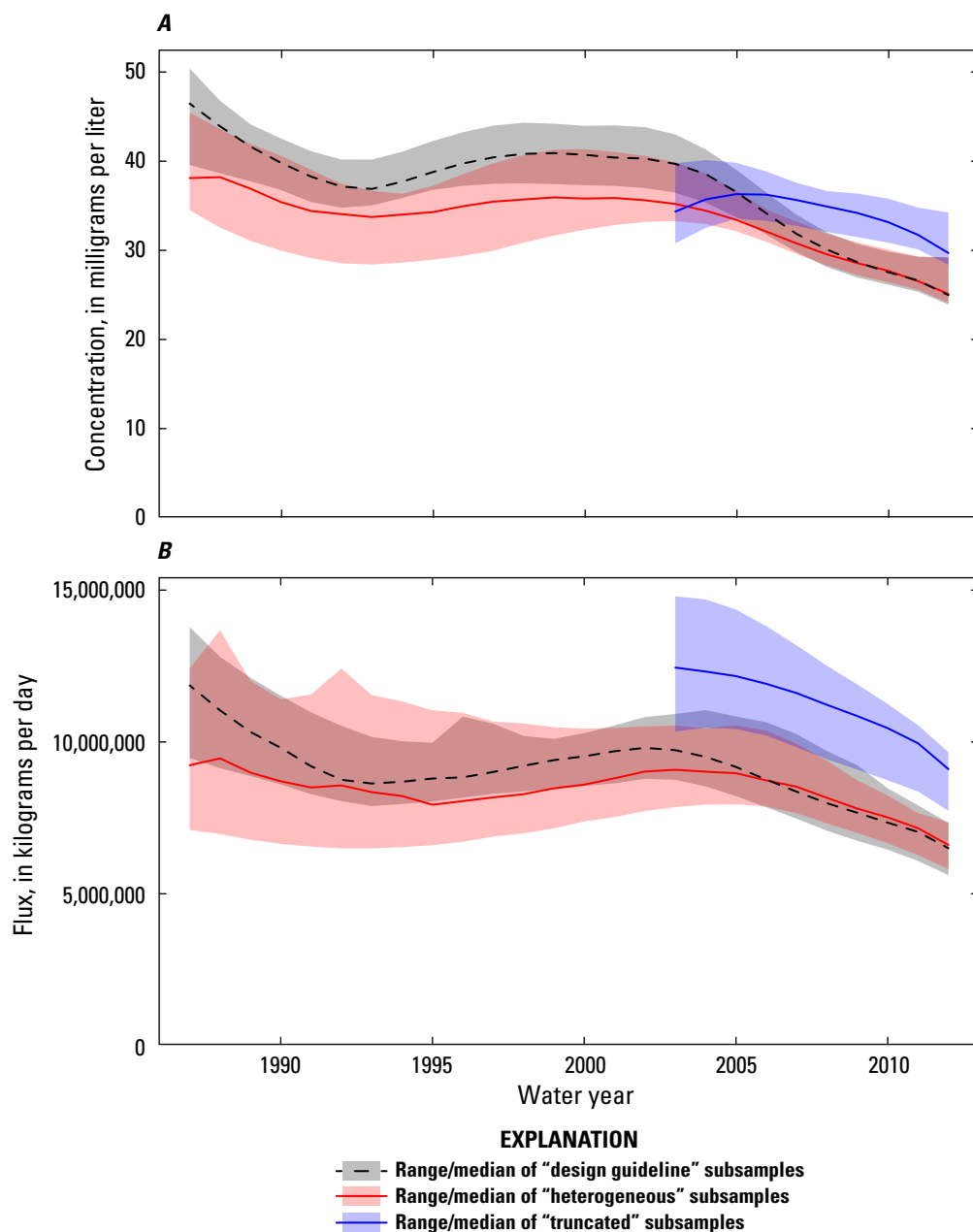


Figure 32. Long-term trends in flow-normalized concentration and flux of sediment at Susquehanna River at Marietta, PA, estimated using 10 “design guideline” subsamples, 10 “heterogeneous” subsamples, and 10 subsamples with the simulation period truncated to 2003–12. *A*, Flow-normalized concentration. *B*, Flow-normalized flux.

as “truncated.” The results of truncating the subsamples are shown in blue lines (year-to-year medians of estimated flow-normalized concentrations for the 10 subsamples) and blue-shaded bands (year-to-year range) in figures 30–32 for dissolved inorganic nitrogen, total phosphorus, and sediment, respectively, at the Susquehanna River at Marietta. The absence of uncertainty bounds on the trends rules out statistical hypothesis tests, but the differences are qualitatively pronounced. For dissolved inorganic nitrogen, the 10-year trend line based on “truncated” subsamples is shifted upward and sloped more steeply downward relative to the corresponding 10-year segment of the long-term trend line estimated using either the design guideline or heterogeneous subsamples (fig. 30). There is less evidence of a change in trend slope for total phosphorus (fig. 31), but an appreciable upward bias remains evident, particularly for the flux trend (fig. 31*B*). Results for sediment (fig. 32) show effects of both positive bias and distortion of slope. Trends based on the “truncated” subsamples for the other seven stations, corresponding to those illustrated by the blue lines and bands in figures 30–32 for the Susquehanna River at Marietta, are included in appendix 5.

The magnitude and consistency of changes in 10-year (2003–12) trend slope for dissolved inorganic nitrogen, total phosphorus, and sediment for the eight stations used in the subsampling experiment are summarized in tables 5–7. For dissolved inorganic nitrogen at the Susquehanna River at Marietta, PA (SRMA, USGS station 01576000), the slope of the 10-year trend line composed of the year-to-year medians of the estimated flow-normalized concentration based on the 10 design guideline subsamples (the segment of the solid black line spanning the years 2003–12 in figure 30*A*), is -2.05 percent per year (first column in table 5). The slope of the corresponding trend line composed of the year-to-year medians of the estimated flow-normalized concentration based on the heterogeneous subsamples (the segment of the red line spanning the years 2003–12 in figure 30*A*) is -2.03 percent per year (second column in table 5). The slope based on the heterogeneous subsamples differs from the slope based on design guideline subsamples by only 0.0189 percent per year (third column in table 5), indicating that, for this station and constituent, the effect of missing storm samples prior to 2003 on 2003–12 trends is negligible. In contrast, the slope of the 10-year trend line obtained using “truncated” subsamples (the blue line in figure 30*A*) was -4.14 percent per year (fourth column in table 5), deviating from the slope based on design guideline subsamples by -2.10 percent per year (fifth column in table 5). Summary statistics for all eight stations (the last three rows in table 5) indicate that the median difference in dissolved inorganic nitrogen trend slope associated with using the heterogeneous subsamples, relative to the design guideline subsamples, was -0.0761 percent per year. The direction of the

differences was not consistent across all stations, ranging from -0.257 to $+0.0189$ percent per year. In contrast, truncating the record to the last 10 years resulted in a median change in slope of -1.88 percent relative to the slope estimated using design guideline subsamples; moreover the change was in the direction of more steeply negative slopes at all stations, ranging from -2.79 to -0.762 percent per year. Results were broadly similar for trends in flow-normalized flux (the last five columns in table 5). Although the same degree of consistency in sign was not shown across all eight stations, differences in slope of the 10-year trends for total phosphorus were generally positive, but more so as a result of truncating the record: The median and maximum differences in concentration trend slopes estimated using the heterogeneous subsamples and the design guideline subsamples were 0.269 and 0.524 percent per year, respectively (table 6). However, the median and maximum differences in slopes estimated using the “truncated” subsamples and the heterogeneous subsamples were 0.598 and 3.42 percent per year, respectively. Similar results were evident for sediment (table 7).

In general, the trend line graphs for all eight stations (appendix 5) and slope differences tabulated in tables 5–7 indicate that preserving a heterogeneous record and reporting flux and trend results only for the portion of the record collected under design guideline sampling protocol is preferable to discarding the portion of the record prior to instituting the protocol. The exact mechanisms underlying the appearance of artifacts in WRTDS flux and trend estimates using data collected with inconsistent sampling protocols are not well understood, and the degree to which the same data could distort flux and trend estimates computed using other regression-based methods has not been assessed. The variety of effects observed and the variety in consistency between stations could be attributed to a range of factors, or combination of factors, including the nature of the concentration-discharge relation for a given constituent, the overall flow regime at the station, hydrologic conditions around the time of the heterogeneity, the degree of contrast in protocol that the heterogeneity represents, or the direction, steepness, or degree of regional consistency in the “true” environmental trend. Despite the fact that the records for the eight stations examined in this experiment represent some of the most densely sampled sites in the CBNTN, it is unlikely that the records represent the full range of potential variability in all of these factors. A more thorough understanding of the factors responsible for inducing the artifacts documented in this section might be obtained using either manipulative studies on more densely sampled research datasets such as those used by Hirsch (2014) or potentially through studies on synthetic datasets. Development of “data quality” diagnostics to identify CBNTN water-quality records with the greatest potential to produce artifacts in WRTDS load and trend estimates is a topic of active research.

Table 5. Ten-year (2003–12) trend slopes for flow-normalized concentration and flux of dissolved inorganic nitrogen for eight Chesapeake Bay Nontidal Water-Quality Monitoring Network sites, comparing the slopes computed using both “heterogeneous” subsamples and “truncated” subsamples with slopes estimated on the basis of “design guideline” subsamples.

Station abbreviation	Station ID	Concentration trends ^{1,2}					Flux trends ^{1,2}				
		Design guideline	Heterogeneous	HT – DG ³	Truncated	TR – DG ⁴	Design guideline	Heterogeneous	HT – DG ³	Truncated	TR – DG ⁴
POTC	01646580	–1.51	–1.49	0.0150	–2.74	–1.24	–1.38	–1.33	0.0542	–2.29	–0.912
SRDA	01540500	–1.72	–1.72	0.00175	–2.48	–0.762	–1.72	–1.65	0.0693	–2.23	–0.514
SRMA	01576000	–2.05	–2.03	0.0189	–4.14	–2.10	–2.14	–2.30	–0.153	–3.61	–1.47
WBSL	01553500	–0.962	–1.00	–0.0408	–2.80	–1.84	–1.80	–1.85	–0.0587	–3.20	–1.41
CONE	01576754	–0.545	–0.656	–0.111	–2.45	–1.91	–0.836	–1.01	–0.170	–2.45	–1.61
SRCW	01578310	–0.764	–0.944	–0.180	–2.91	–2.14	–1.05	–1.32	–0.273	–2.83	–1.78
APPM	02041650	1.35	1.22	–0.129	–0.210	–1.56	–2.25	–2.50	–0.242	–3.83	–1.58
MABU	01674500	–0.320	–0.577	–0.257	–3.11	–2.79	–0.676	–1.04	–0.364	–3.38	–2.70
Minimum		–2.05	–2.03	–0.257	–4.14	–2.79	–2.25	–2.50	–0.364	–3.83	–2.70
Median		–0.863	–0.973	–0.0761	–2.77	–1.88	–1.55	–1.49	–0.162	–3.01	–1.52
Maximum		1.35	1.22	0.0189	–0.210	–0.762	–0.676	–1.01	0.0693	–2.23	–0.514

¹All values characterize the trend line consisting of the year-to-year medians of annual values for 10 subsamples.

²All results and comparisons are in average percent change per year over the period 2003–12, where total change is calculated as a percentage of the value in 2003.

³Difference in 10-year trend between “heterogeneous” subsamples and “design guideline” subsamples.

⁴Difference in 10-year trend between “truncated” subsamples and “design guideline” subsamples.

Table 6. Ten-year (2003–12) trend slopes for flow-normalized concentration and flux of total phosphorus for eight Chesapeake Bay Nontidal Water-Quality Monitoring Network sites, comparing the slopes computed using both “heterogeneous” subsamples and “truncated” subsamples with slopes estimated on the basis of “design guideline” subsamples.

Station abbreviation	Station ID	Concentration trends ^{1,2}				Flux trends ^{1,2}					
		Design guideline	Hetero- geneous	HT – DG ³	Truncated	TR – DG ⁴	Design guideline	Hetero- geneous	HT – DG ³	Truncated	TR – DG ⁴
POTC	01646580	-2.60	-2.08	0.524	0.821	3.42	-1.59	-0.521	1.06	0.847	2.43
SRDA	01540500	-3.52	-3.51	0.0112	-3.19	0.325	-3.10	-3.08	0.0209	-2.90	0.205
SRMA	01576000	-3.73	-3.36	0.373	-2.86	0.871	-2.89	-2.71	0.175	-2.71	0.173
WBSL	01553500	-5.56	-5.37	0.194	-5.42	0.139	-5.03	-4.74	0.293	-4.56	0.473
CONE	01576754	-2.38	-1.88	0.496	0.957	3.34	-1.26	-0.593	0.665	0.121	1.38
SRCW	01578310	0.0553	0.399	0.343	0.254	0.199	1.77	2.52	0.747	2.39	0.618
APPM	02041650	0.585	0.441	-0.145	0.543	-0.0420	0.380	0.216	-0.163	0.851	0.472
MABU	01674500	0.766	0.900	0.133	1.66	0.896	1.26	1.63	0.362	2.24	0.975
Minimum		-5.56	-5.37	-0.145	-5.42	-0.0420	-5.03	-4.74	-0.163	-4.56	0.173
Median		-2.49	-1.98	0.269	0.399	0.598	-1.42	-0.557	0.327	0.484	0.546
Maximum		0.766	0.900	0.524	1.66	3.42	1.77	2.52	1.06	2.39	2.43

¹All values characterize the trend line consisting of the year-to-year medians of annual values for 10 subsamples.

²All results and comparisons are in average percent change per year over the period 2003–12, where total change is calculated as a percentage of the value in 2003.

³Difference in 10-year trend between “heterogeneous” subsamples and “design guideline” subsamples.

⁴Difference in 10-year trend between “truncated” subsamples and “design guideline” subsamples.

Table 7. Ten-year (2003–12) trend slopes for flow-normalized concentration and flux of sediment for eight Chesapeake Bay Nontidal Water-Quality Monitoring Network sites, comparing the slopes computed using both “heterogeneous” subsamples and “truncated” subsamples with slopes estimated on the basis of “design guideline” subsamples.

Station abbreviation	Station ID	Concentration trends ^{1,2}				Flux trends ^{1,2}					
		Design guideline	Hetero- geneous	HT – DG ³	Truncated	TR – DG ⁴	Design guideline	Hetero- geneous	HT – DG ³	Truncated	TR – DG ⁴
POTC	01646580	0.454	0.419	−0.0348	2.93	2.48	0.967	1.15	0.180	0.777	−0.191
SRDA	01540500	−2.03	−1.38	0.648	0.321	2.35	−1.09	0.0989	1.19	1.39	2.48
SRMA	01576000	−3.71	−2.87	0.844	−1.35	2.36	−3.32	−2.73	0.592	−2.69	0.633
WBSL	01553500	−4.09	−3.10	0.992	−3.14	0.953	−4.84	−3.39	1.45	−3.37	1.47
CONE	01576754	−2.70	−1.80	0.903	0.572	3.28	−2.66	−1.52	1.14	−3.02	−0.365
SRCW	01578310	0.867	1.19	0.324	3.07	2.21	0.454	1.09	0.633	2.06	1.60
APPM	02041650	1.80	1.93	0.127	3.16	1.36	0.487	0.716	0.229	1.73	1.25
MABU	01674500	−0.389	0.664	1.05	0.981	1.37	−0.734	0.361	1.09	0.271	1.01
Minimum		−4.09	−3.10	−0.0348	−3.14	0.953	−4.84	−3.39	0.180	−3.37	−0.365
Median		−1.21	−0.481	0.746	0.777	2.28	−0.913	0.230	0.864	0.524	1.13
Maximum		1.80	1.93	1.05	3.16	3.28	0.967	1.15	1.45	2.06	2.48

¹All values characterize the trend line consisting of the year-to-year medians of annual values for 10 subsamples.

²All trends reflect mean annual concentration/flux in 2012 minus corresponding value in 2003, expressed as a percentage of the value in 2003.

³Difference in 10-year trend between “heterogeneous” subsamples and “design guideline” subsamples.

⁴Difference in 10-year trend between “truncated” subsamples and “design guideline” subsamples.

Changes in Laboratory Methods

With ongoing advances in analytical chemistry, it is inevitable that long-term records of concentration for many constituents reflect occasional changes in laboratory methods. Ideally, changes in methods would be completely transparent to subsequent examination or analysis; realistically, however, such changes are often evident. Some of the more readily detectable results of method changes include changes in rounding convention, changes in censoring frequency, and changes in censoring level; such changes are often evident upon close visual examination of plots of concentration in relation to time. Less evident are subtle biases that are not accompanied by any visual cues. Although some artifacts can be handled with modern statistical techniques, others reflect historical methodological limitations the effects of which may never be completely resolved. The latter can lead to difficult choices as to how the data should be interpreted.

Examples of the more evident changes in laboratory methods are shown in figure 33. The first example shows an obvious change in orthophosphate reporting convention, presumably associated with a change in laboratory methods, around 1994. The change includes the removal of a censoring threshold at 0.01 mg/L, evident as the disappearance of vertical lines indicating censoring, and a change in rounding convention, evident as the clustering of observations at discrete ordinates above 0.01 mg/L prior to 1994 and their dispersed distribution above 0.01 mg/L thereafter. The second example shows a shift in censoring level from predominately 0.002 mg/L until about 1996 up to 0.01 mg/L after around 2002. Censoring in environmental data has been the subject of detailed study over the past several decades (for example, Helsel, 2012), and defensible approaches for summarizing, correlating, and performing linear regression with censored data are well established. Overall, the proportions of total observations in the first and second examples that are censored in figure 33 are 10.2 and 20.2 percent, respectively. Survival regression with maximum likelihood parameter estimation, the method used by WRTDS, has been demonstrated to produce credible results on a water-quality dataset with an overall censoring rate in excess of 50 percent (Helsel, 2012). The default upper limit on the proportion of censored observations in any given call to the survival regression routine in WRTDS is 50 percent. Thus, even for the long-term dataset with the highest overall proportion of censored samples (44.5 percent for total suspended solids at the North Fork Shenandoah River near Strasburg, VA [NFSH, USGS station 01634000]; data not shown), the censoring frequency is well below the limit of documented successful application of survival regression.

A more relevant question for data such as those shown in figure 33 is the extent to which monotonic sequential steps in censoring level have the potential to confound detection and quantification of trends. A preliminary exploration of this potential was conducted using the record for total phosphorus at the Conestoga River at Conestoga, PA (CONE, USGS station 01576754), from water years 1985 to 2012 inclusive

(fig. 34). The record of discrete observations, none of which are censored, conveys a general impression of a downward trend in concentration (fig. 34A); the period-of-record trend in flow-normalized concentration is -68.4 percent (appendix 2). Observations were grouped into three non-overlapping time periods, each containing one-third of the total observation count. Within these groups, censoring levels were artificially imposed at concentrations of 0.1, 0.2, and 0.5 mg/L in two “stair-step” patterns representing both increasing (fig. 34B) and decreasing (fig. 34C) levels over time; overall, this manipulation resulted in censoring frequencies of 39.7 and 41.2 percent in the “increasing” and “decreasing” scenarios, respectively. Trends in flow-normalized concentration of total phosphorus resulting from running WRTDS on these datasets (-63.6 percent for the “decreasing” scenario and -70.3 percent for the “increasing” scenario) are shown in figure 35. Although there is some variation between the three trend lines, the trends based on the two censored datasets represented the overall downward trend estimated from the “baseline” data rather closely; the pattern of variation in trends in flow-normalized flux (not shown) was nearly identical to that shown in figure 35 for concentration. In general, the sensitivity of a trend result to changes in censoring level could be expected to be related to a variety of factors, including the strength of the “true” trend, the overall range of the data, the overall frequency of censored observations, and the separation in the different censoring levels. The results of this single exploratory exercise do not rule out the possibility that censoring issues might influence trend results in some cases but lend a degree of confidence that the survival regression routines upon which WRTDS are based are generally robust to the types of censoring issues evident in CBNTN data.

An example of a more subtle impact of a change in laboratory methods is evident in the record for total phosphorus at the Choptank River near Greensboro, MD (CHOP, USGS station 01491000; fig. 36). In January 1999, the National Water Quality Laboratory (NWQL) switched analytical methods for total phosphorus, including both new analysis methods (U.S. Geological Survey, 1999a) and sample preservation methods (U.S. Geological Survey, 1999b). Associated with the time of the NWQL change (as well as a change in rounding convention) is a step shift upward in the lower concentration range in figure 36. Superimposed on the concentration data, in green, is the period-of-record trend in flow-normalized concentration estimated using all the concentration data. The trend line turns upward beginning around 1995. Note that because the time windowing feature of WRTDS is essentially a smoothing mechanism, WRTDS is not an appropriate tool for detecting step-changes in trend estimates. In the current example, because time windowing includes observations both before and after the time for which an estimate is required in its weighting scheme, it can be expected that the first sign of an effect, on estimated trends, of an abrupt discontinuity in observed data will precede the discontinuity. One “null hypothesis” to explain the upward slant in the trend line is that it reflects the influence of four

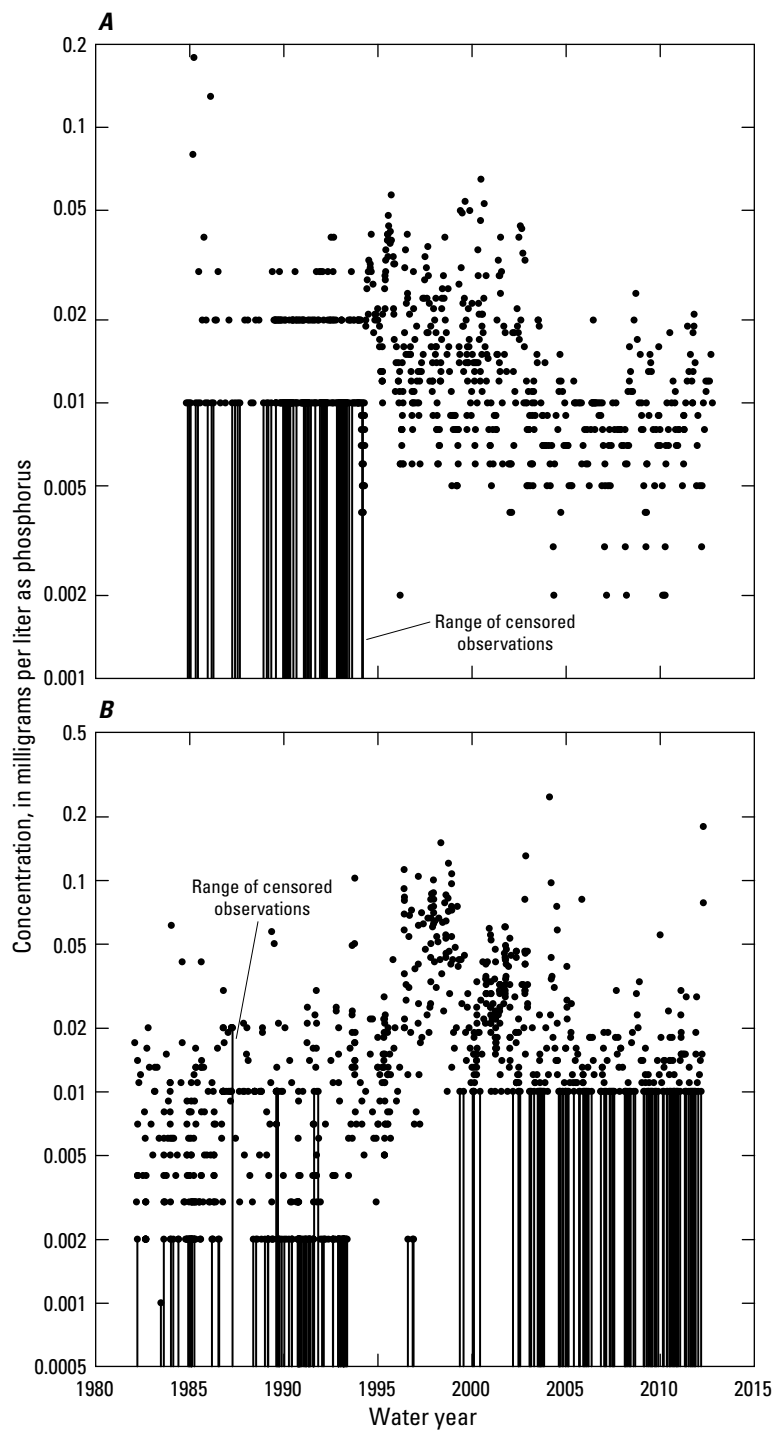


Figure 33. Plots of observed concentration of orthophosphate over time at two long-term stations, illustrating inconsistent censoring. *A*, Mattaponi River near Beulahville, VA, showing a shift in the frequency of censored observations. *B*, Susquehanna River at Marietta, PA, showing a step change in the censoring level.

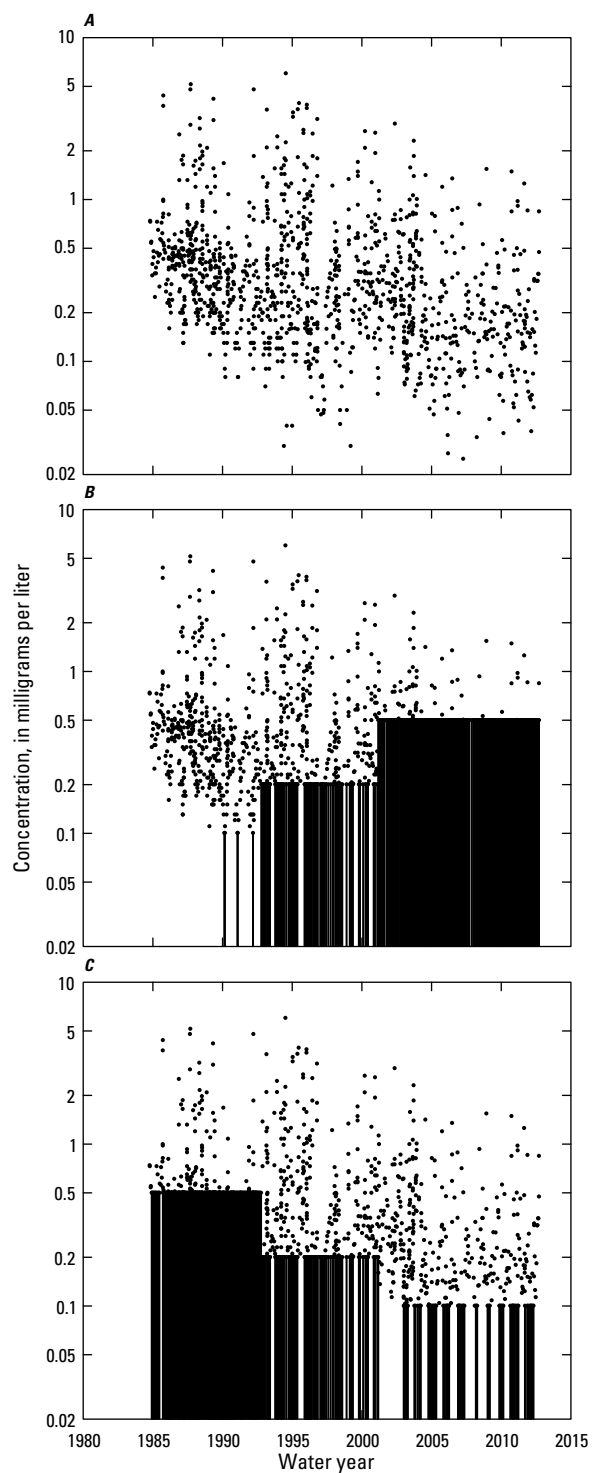


Figure 34. Plots of observed concentration of total phosphorus over time at the Conestoga River at Conestoga, PA, illustrating artificially imposed censoring conditions. *A*, Original data. *B*, Imposed increasing censoring threshold. *C*, Imposed decreasing censoring threshold.

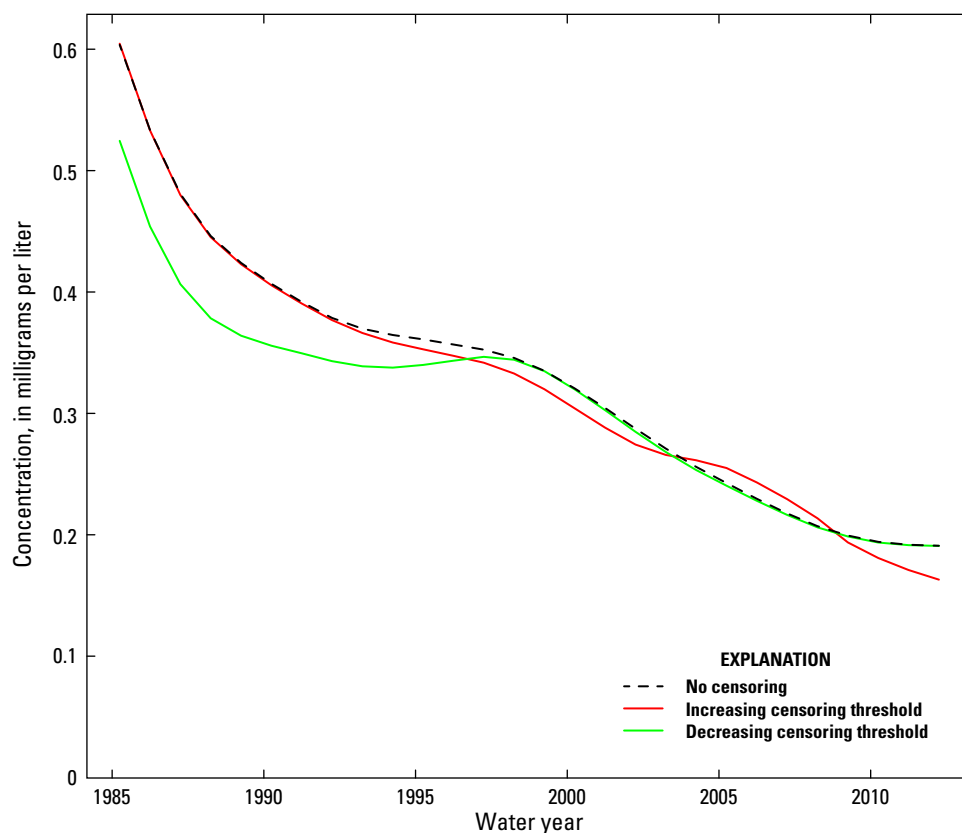


Figure 35. Trends in flow-normalized concentration of total phosphorus at the Conestoga River at Conestoga, PA, contrasting trends estimated using the three datasets illustrated in figure 34.

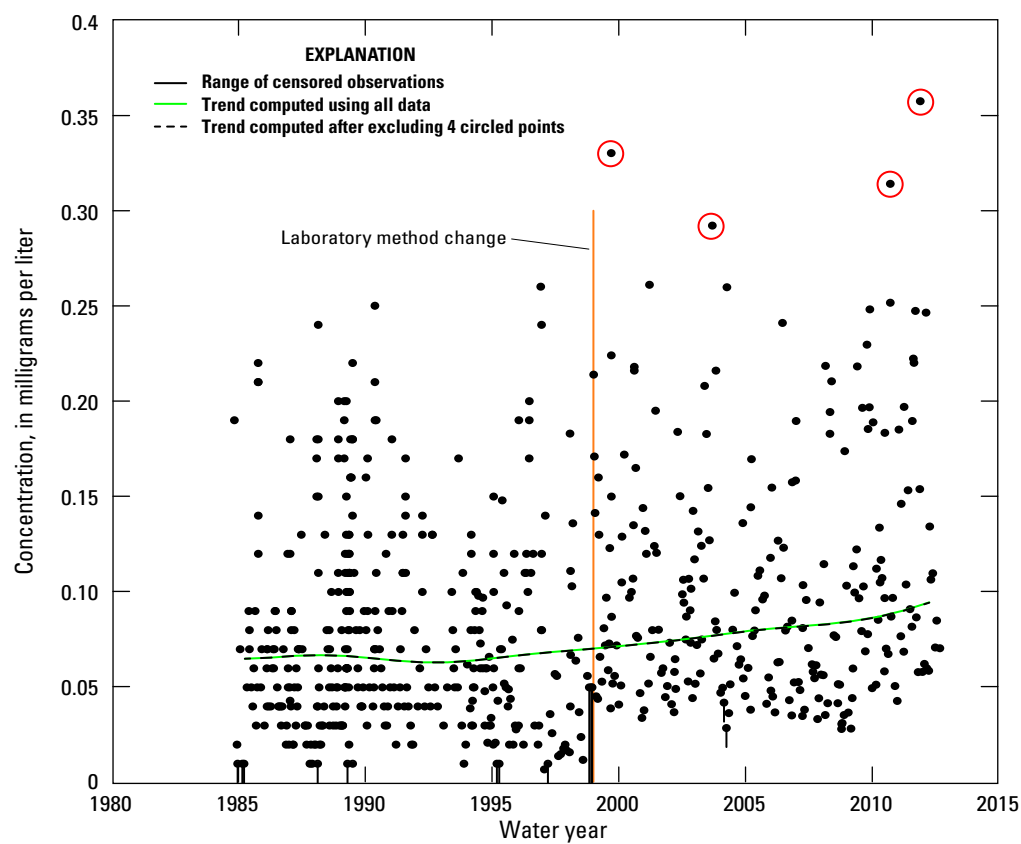


Figure 36. Plot of observed concentration of total phosphorus over time at the Choptank River near Greensboro, MD, with superimposed trends in flow-normalized concentration, illustrating the effects of a change in laboratory methods in January 1999.

relatively high-concentration observations (circled in red in figure 36), observed beginning in 1999. However, the trend line computed using all but the four circled observations, shown in figure 36 as the dashed black line, is nearly identical to the trend on the basis of all the data (for example, the 2012 annual flow-normalized concentrations computed based on the full and reduced datasets are 0.0940 and 0.0941 mg/L, respectively). The most evident alternative is that the upward shift is due to some combination of (a) the abrupt disappearance of all concentration values below about 0.04 mg/L after 1999 and (b) any potentially less-easy-to-observe shifts in data in other concentration ranges beginning around the same time. Either a or b could reflect some combination of the effects of the 1999 laboratory methods changes and “true” changes in watershed processes. The fact that similar phenomena associated with the 1999 total phosphorus method changes have been observed nationwide in samples processed by the NWQL (Lori Sprague, U.S. Geological Survey, oral commun., September 11, 2014) supports the possibility that the upward shift in the trend line is at least in part an artifact of the method change. Methods for isolating the effects of the 1999 total phosphorus method change, and method changes in general, are actively being researched. Nevertheless, it remains unclear as to whether any type of generally applicable corrective measure after the 1999 method change will be identified.

Another example potentially relevant to the CBNTN is a switch from Kjeldahl digestion to alkaline persulfate digestion for the determination of total nitrogen. In both synthetic and field samples, the former method has been found to be positively biased in the presence of excess nitrate, and the latter method has been found to be negatively biased in the presence of suspended sediment (Rus and others, 2012). The magnitude of the negative bias in field samples was station specific and dependent on suspended-sediment concentration; thus, no analysis to determine a post-1999 method change correction was undertaken.

Beyond methods research and development, one logical method for verifying that a shift in laboratory methods has not introduced a bias in a water-quality record is to split environmental samples and process them using both methods for a certain period of time until a certain number of samples has been collected or until some specified range of environmental conditions has been encountered. Such procedures, however, are not always feasible within the budget limitations of smaller projects or jurisdictional monitoring programs, which frequently are important sources of data for constructing long-term records. Moreover, without careful statistical design, including assumptions about the potential magnitude of methods-related change and conditions under which they occur, there is no guarantee that a specified period of overlapping methods will completely characterize the extent of any problem.

In the absence of a credible quantitative remedy to the problem of “step” trends associated with methods changes, the two most obvious means of dealing with these issues are to (a) only include periods with homogeneous methods in the

trend analysis, or (b) analyze the entire record and qualify the results. Truncating the record to the most recent period of consistent methods completely eliminates the question of bias. However, as was demonstrated previously for inconsistent storm sampling, simply truncating the record, at least to 10 years, can in some cases induce artifacts that could influence conclusions about overall trend direction. On the other hand, a small methods-related bias in a certain concentration range, or under certain environmental conditions, might not be considered sufficient grounds for discarding a large portion of the record, often representing a substantial investment of resources; it could be argued that data collected under historical methods that are anywhere short of being widely discredited still contain a good degree of useful information under some range of environmental conditions. The obvious issue with this argument is that the exact range of conditions under which the historical data can be considered reliable are seldom known.

For the determination of fluxes and trends in the CBNTN, USGS Bay Program staff will make decisions related to truncating records to periods of consistent laboratory methods on a case-by-case basis. As a rule, long-term records should be preserved in deference to the effort of the cooperators who committed resources to ongoing environmental monitoring. However, any analyses that are based on a long-term record that includes a method change with documented potential to influence flux or trend results should be qualified, and a log of such method changes and affected records should be maintained. Stations that appear to warrant more detailed study, including potential re-analysis with a truncated record, can be considered individually as dictated by the need to better understand station-specific results.

Summary

In their water year 2012 online posting of fluxes and trends of nutrients and sediment from the Chesapeake Bay watershed into the estuary, the U.S. Geological Survey (USGS) Chesapeake Bay Program staff adopted Weighted Regressions on Time, Discharge, and Season (WRTDS) as a tool for reporting fluxes, but retained ESTIMATOR as a tool for reporting trends, primarily on the basis of the novelty of the “flow-normalization” algorithm employed by WRTDS. The first part of this report documents a detailed comparison of the two models, with a particular focus on trends. The key findings of that comparison can be summarized as follows:

1. Application of WRTDS to estimate fluxes across a wide range of records from the Chesapeake Bay Nontidal Water-Quality Monitoring Network (CBNTN) demonstrated that the same advantages over ESTIMATOR, reported by other researchers who used carefully selected water-quality records, are also evident in the broader setting of an operational water-quality network. For the 30 stations having record lengths

exceeding 25 years, median model explanatory power (approximately, R^2) for WRTDS improved by 6 percent for total nitrogen (33.9 to 39.9 percent), 8 percent for total phosphorus (38.6 to 46.9 percent), and 5 percent for sediment (56.5 to 61.2 percent), relative to ESTIMATOR; smaller improvements were noted with shorter records. Flux bias statistics for WRTDS decreased (improved) relative to ESTIMATOR, most notably for sediment (+0.249 to +0.087; about 17 percentage points for stations with record lengths exceeding 25 years) and total phosphorus (+0.047 to +0.008; about 5 percentage points); a similar degree of improvement was evident with shorter records.

2. WRTDS estimates of flow-normalized trend in concentration generally track ESTIMATOR flow-adjusted trend in concentration, in magnitude and direction, for stations having record lengths exceeding 25 years; the correspondence became much closer for stations with 10-year records.
3. This correspondence can be traced to the operational similarities in the models' trend estimation approaches, despite conceptual differences. In light of these similarities, the flow-normalization approach used in WRTDS can be thought of as a more flexible extension of that used by ESTIMATOR rather than a new, unrelated approach.
4. By virtue of that flexibility, WRTDS was demonstrated to produce more credible estimates of long-term concentration trends at the nine RIM stations than ESTIMATOR.
5. Whereas ESTIMATOR is constrained to reporting trends in concentration only, the model structure of WRTDS allows for independent estimation of trends in concentration and flux; because trends in fluxes and trends in concentration reflect the effects of different watershed processes, the ability to estimate trends in fluxes represents an advantage for interpretation.
6. Trends estimated by both ESTIMATOR and WRTDS showed a similar degree of sensitivity to incremental addition of new data.

On the basis of these findings, USGS Bay Program staff are adopting WRTDS as the primary tool for future reporting of both flux and trend results for the CBNTN effective on the next reporting cycle, currently projected to include data through water year 2014. Uncertainty bounds, not available with the version of WRTDS used to produce water year 2012 fluxes and trends reported herein, are expected to be available for water year 2014 products.

Nutrient and sediment flux and trend estimates, based on WRTDS, for all stations for which fluxes or trends were reported in water year 2012, are summarized narratively in the second part of this report and are tabulated in appendixes 2

(period-of-record and 10-year trends) and 3 (annual time series); these constitute the water year 2012 results of record for scientific analysis and regulatory interpretation.

The CBNTN reflects one of the first applications of WRTDS, to date, predominantly an exploratory research tool, in a large operational water-quality network. In the third part of this report, some of the potential effects of inconsistencies in sampling and laboratory analysis protocols on modeled flux and trend results were explored, including features common in many large-scale monitoring networks. Three "non-ideal" features evident in some of the CBNTN records include inconsistency in annual sampling effort, inconsistency in storm sampling effort, and changes in laboratory methods. It is conceivable that flux or trend estimates based on such records may in some cases reflect the effects of these heterogeneities rather than true trends in environmental conditions. The potential effects of the first two of these heterogeneities were explored, using WRTDS, by conducting a subsampling experiment using eight of the most densely sampled long-term (>25 years) stations in the CBNTN as baseline datasets.

From each dataset, a set of 10 "design guideline" subsamples consisting of 20 samples per year was drawn in a manner that preserved the overall intensity of storm sampling in the baseline data. These 10 subsamples were further manipulated to create heterogeneous subsamples, mimicking the effect of introducing storm sampling in the last 10 years of a record that began with only monthly sampling. Across the eight stations, the maximum relative difference between flow-normalized flux estimated in a single year from the 10 design guideline subsamples and values estimated in the corresponding year from baseline data was smallest for dissolved inorganic nitrogen (median of 8 stations = 6 percent of baseline estimate), but more appreciable for total phosphorus (typically predominantly particulate) and sediment (medians of 22 and 32 percent, respectively). Despite this degree of variability, broad patterns in period-of-record trends based on the subsamples closely resembled the trend estimated from baseline data; statistical hypothesis testing was not possible given the lack of uncertainty bounds on long-term trends. The maximum relative difference between flow-normalized flux estimated from the 10 heterogeneous subsamples and values estimated in the corresponding year from baseline data was more pronounced, with medians for 8 stations of 15, 30, and 53 percent of the corresponding baseline estimates for dissolved inorganic nitrogen, total phosphorus, and sediment, respectively. The worst-case maximum relative differences between flow-normalized flux estimated in a single year from the 10 heterogeneous subsamples and values estimated in the corresponding year from baseline data were 25 percent for dissolved inorganic nitrogen, 37 percent for total phosphorus, and 250 percent for sediment. The results for the heterogeneous subsamples indicate that changes in high-flow sampling protocol can result in appreciable distortion of estimated trends in flow-normalized flux, especially for phosphorus and sediment. Trend lines estimated from heterogeneous subsamples tended to converge with the trend

lines estimated from baseline data after 2003. Truncating the record by discarding the data prior to the 2003 heterogeneity, an approach sometimes advocated, resulted in changes in trend slope that were frequently larger than those seen between the design guideline and heterogeneous subsamples. For nitrate plus nitrite, trend slopes for flow-normalized concentration over the last 10 years of the record were consistently more steeply negative, relative to the design guideline scenario, when the record was truncated to 10 years (median difference in slope of -1.88 percent per year), than was the case when the complete but heterogeneous record was retained (median difference in slope of -0.0761 percent per year); similar results were evident for flow-normalized flux. Similar but less-pronounced results were evident for total phosphorus and sediment. These quantitative differences are probably related to unique features of these records. Overall, the results indicate that for particulate constituents, load and trend estimates computed using long-term records recently converted to CBNTN design guideline sampling protocols will be most reliable if the trend is computed using the entire record but reported only for the period that design guideline sampling protocols were followed.

Two examples of heterogeneities related to changing laboratory methods were examined by conducting simple manipulative experiments. The first example considered the effects of changes in laboratory censoring levels over time, using a record for total phosphorus having a rather steep (-68.4 percent over 27 years) downward trend and (initially) no censored observations. Two “stair-step” changes in censoring level, overall representing a factor-of-five change in the laboratory reporting limit, were artificially imposed: one reflecting increasing levels over time and the other decreasing levels. Trends estimated on the basis of these rather heavily manipulated datasets were broadly similar to the original trend (-63.6 percent for decreasing censoring levels and -70.3 percent for increasing censoring levels), lending a degree of confidence that the survival regression routines upon which WRTDS is based are generally robust to data censoring. The second example considered an abrupt disappearance of low-concentration (less than about 0.04 milligrams per liter) observations of total phosphorus, associated with a laboratory method change and not reflected through censoring, near the middle of a 28-year record. By process of elimination, an upward shift in the estimated flow-normalized concentration trend line around the same time was identified as a likely artifact resulting from the laboratory method change, although a contemporaneous change in watershed processes cannot be ruled out. Decisions as to how to treat records with potential methods-related artifacts should be made on a case-by-case basis, and trend results should be appropriately qualified.

References Cited

- Chesapeake Bay Program Office, 2009, Recommendations to improve coordinated nontidal monitoring, assessment, and communication activities in support of Chesapeake Bay restoration—A report addressing STAC recommendation for monitoring reallocation, 2009: Annapolis, Md., Chesapeake Bay Program, 40 p., accessed June 24, 2015, at http://www.chesapeakebay.net/channel_files/21466/recommendations_to_improve_nontidal_monitoring_a_report_addressing_stac_recommendations_2009.pdf.
- Chesapeake Bay Program Office [2014], Learn the issues: Chesapeake Bay Program Web page, accessed December 21, 2014, at <http://www.chesapeakebay.net/issues>.
- Claggett, Peter, and Thompson, Renee, eds., 2012, Proceedings of the workshop on alternative futures—Accounting for growth in the Chesapeake Bay watershed: U.S. Geological Survey Open-File Report 2012–1216, 29 p.
- Cohn, T.A., Caulder, D.L., Gilroy, E.J., Zynjuk, L.D., and Summers, R.M., 1992, The validity of a simple statistical model for estimating fluvial constituent loads—An empirical study involving nutrient loads entering Chesapeake Bay: *Water Resources Research*, v. 28, no. 9, p. 2353–2363.
- Cohn, T.A., Delong, L.L., Gilroy, E.J., Hirsch, R.M., and Wells, D.K., 1989, Estimating constituent loads: *Water Resources Research*, v. 25, no. 5, p. 937–942.
- Corsi, S.R., De Cicco, L.A., Lutz, M.A., and Hirsch, R.M., 2015, River chloride trends in snow-affected urban watersheds—Increasing concentrations outpace urban growth rate and are common among all seasons: *Science of the Total Environment*, v. 508, p. 488–497.
- Gilroy, E.J., Hirsch, R.M., and Cohn, T.A., 1990, Mean-square error of regression-based constituent transport estimates: *Water Resources Research*, v. 26, no. 9, p. 2069–2077.
- Glysson, G.D., Gray, J.R., and Schwarz, G.E., 2001, A comparison of load estimates using total suspended solids and suspended-sediment concentration data, in Phelps, D., and Sehlke, G., eds., *Bridging the gap—Meeting the world’s water and environmental resources challenges*, Proceedings of the World Water and Environmental Resources Congress, Orlando, Fla., May 20–24, 2001, p. 1–10.
- Green, C.T., Bekins, B.A., Kalkhoff, S.J., Hirsch, R.M., Liao, L., and Barnes, K.K., 2014, Decadal surface water quality trends under variable climate, land use, and hydrogeochemical setting in Iowa, U.S.A.: *Water Resources Research*, v. 50, no. 3, p. 2425–2443.
- Helsel, D.R., 2012, *Statistics for censored environmental data using Minitab and R* (2d ed.): New York, John Wiley and Sons, 344 p.

- Helsel, D.R., and Hirsch, R.M., 2002, Statistical methods in water resources: U.S. Geological Survey Techniques of Water-Resources Investigations, book 4, chap. A3, 510 p.
- Hirsch, R.M., 2014, Large biases in regression-based constituent flux estimates—Causes and diagnostic tools: *Journal of the American Water Resources Association*, v. 50, no. 6, p. 1401–1424.
- Hirsch, R.M., and De Cicco, Laura, 2014, User guide to exploration and graphic for RivEr Trends (EGRET) and dataRetrieval—R packages for hydrologic data: U.S. Geological Survey Techniques and Methods, book 4, chap. A10, 95 p.
- Hirsch, R.M., Moyer, D.L., and Archfield, S.A., 2010, Weighted regressions on time, discharge, and season (WRTDS), with an application to Chesapeake Bay river inputs: *Journal of the American Water Resources Association*, v. 46, no. 5, p. 857–880.
- Langland, M.J., Blomquist, J.D., Moyer, D.L., Hyer, K.E., and Chanut, J.G., 2013, Total nutrient and sediment loads, trends, yields, and nontidal water-quality indicators for selected nontidal stations, Chesapeake Bay Watershed, 1985–2011: U.S. Geological Survey Open-File Report 2013–1052, 51 p.
- Langland, M.J., Lietman, P.L., and Hoffman, S.A., 1995, Synthesis of nutrient and sediment data for watersheds within the Chesapeake Bay drainage basin: U.S. Geological Survey Water-Resources Investigations Report 95–4233, 121 p.
- Langland, M.J., Raffensperger, J.P., Moyer, D.L., Landwehr, J.M., and Schwarz, G.E., 2006, Changes in streamflow and water quality in selected nontidal basins in the Chesapeake Bay watershed, 1985–2004: U.S. Geological Survey Scientific Investigations Report 2006–5178, 75 p. 8 apps. (on CD).
- Moyer, D.L., Hirsch, R.M., and Hyer, K.E., 2012, Comparison of two regression-based approaches for determining nutrient and sediment fluxes and trends in the Chesapeake Bay watershed: U.S. Geological Survey Scientific Investigations Report 2012–5244, 118 p.
- Murphy, J.C., Hirsch, R.M., and Sprague, L.A., 2014, Antecedent flow conditions and nitrate concentrations in the Mississippi River basin: *Hydrology and Earth System Sciences*, v. 18, no. 3, p. 967–979.
- Obama, Barack, 2009, Chesapeake Bay protection and restoration: *Federal Register*, v. 74, no. 93, May 15, p. 23099–23104. [Also available at <https://www.gpo.gov/fdsys/pkg/FR-2009-05-15/pdf/E9-11547.pdf>.]
- R Core Team, 2013, R—A language and environment for statistical computing: Vienna, Austria, R Foundation for Statistical Computing, accessed June 25, 2015, at <http://www.R-project.org/>.
- 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.
- Rus, D.L., Patton, C.J., Mueller, D.K., and Crawford, C.G., 2012, Assessing total nitrogen in surface-water samples—Precision and bias of analytical and computational methods: U.S. Geological Survey Scientific Investigations Report 2012–5281, 38 p.
- Sprague, L.A., 2001, Effects of storm-sampling frequency on estimation of water-quality loads and trends in two tributaries to Chesapeake Bay in Virginia: U.S. Geological Survey Water-Resources Investigations Report 2001–4136, 38 p.
- Sprague, L.A., Hirsch, R.M., and Aulenbach, B.T., 2011, Nitrate in the Mississippi River and its tributaries, 1980 to 2008—Are we making progress?: *Environmental Science & Technology*, v. 45, no. 17, p. 7209–7216.
- U.S. Environmental Protection Agency, 2010, Executive summary of the final Chesapeake Bay Total Maximum Daily Load (TMDL): U.S. Environmental Protection Agency Web page, accessed December 21, 2014, at <http://www.epa.gov/reg3wapd/tmdl/ChesapeakeBay/tmdlexec.html>.
- U.S. Geological Survey, 1999a, Validation of U.S. Environmental Protection Agency (USEPA) method 365.1 to replace U.S. Geological Survey (USGS) TWRI methods I-2607-90 and I-4607-90 for determination of phosphorus in the concentration range from 0.004 to 0.200 mg-P/L: National Water Quality Laboratory Technical Memorandum 1999.05, January 7, 1999, 4 p.
- U.S. Geological Survey, 1999b, Changes in field treatment protocols and bottle types for whole-water samples collected for total ammonium plus organic nitrogen and total phosphorus determinations: U.S. Geological Survey Office of Water Quality Technical Memorandum 99.04, November 25, 1998, 4 p.
- U.S. Geological Survey, 2010, USGS science for the Chesapeake Bay restoration: U.S. Geological Survey Fact Sheet 2010–3081, 2 p.
- Zhang, Q., Brady, D.C., and Ball, W.P., 2013, Long-term seasonal trends of nitrogen, phosphorus, and suspended sediment load from the non-tidal Susquehanna River Basin to Chesapeake Bay: *Science of the Total Environment*, v. 452, p. 208–221.

Appendixes

- Appendix 1.** Effects of the incremental annual incorporation of new water-quality observations on trends—Graphic results for the six watersheds subject to detailed study (available at <http://dx.doi.org/10.3133/sir20155133>)
- Appendix 2.** Tabular WRTDS results for period-of-record and 10-year flow-normalized trends in concentration and flux (available at <http://dx.doi.org/10.3133/sir20155133>)
- Appendix 3.** Tabular WRTDS results for all stations and constituents modeled through water year 2012 (available on CD-ROM and at <http://dx.doi.org/10.3133/sir20155133>)

The enclosed CD-ROM contains listings of annual and monthly concentration and flux results, based on models estimated using WRTDS, for all CBNTN stations and constituents for which flux and (or) trend results through water year 2012 were published on the CBNTN Web site (<http://cbrim.er.usgs.gov/index.html>). Findings in this report motivated one modification to WRTDS along with several modifications to its application across the CBNTN that were adopted concurrently with the final preparation of the manuscript; the results tabulated in this appendix were computed with these modifications in place. As such, these results may differ from numerical results discussed or tabulated in the body of the report; however, they reflect the most up-to-date CBNTN results through water year 2012 available at the time this report was published. The specific differences in model formulation and application between results reported in the body of the report and those reported in this appendix are as follows:

1. Partly in response to the findings reported in the section “Sensitivity of Trends to Incremental Incorporation of New Data,” the version of WRTDS used to produce the results tabulated in this appendix, distributed in the EGRET R package v. 2.2.1 and all subsequent versions, was modified to reduce curvature in flow-normalized trends near the beginning and ends of the record. The feature, called “edgeAdjust,” is an input parameter to the routine `modelEstimation()` and is enabled by default. Concurrent with its implementation is a reduction in the default time half-window width from 10 to 7 years.
2. As discussed in the section “Dense Sub-Daily Sampling,” in some instances more than one water-quality observation was recorded on a single day. In producing results reported in the body of the report, only the two observations representing the highest and lowest concentration observed on that day were retained. Subsequently, the model developers recommended that all samples on such days be replaced by a single sample, representing the median of all the samples observed on that day; this recommendation was adopted in producing the results tabulated in this appendix.
3. As described in the section “Description of Scenarios Considered,” the historical practice for estimating 10-year trends at stations where the record length exceeds 10 years has been to truncate the water-quality record to the most recent 10-year period. Results presented in the section “Results: Variability in Storm Sampling Effort” indicate that in some cases 10-year trends obtained from long-term datasets in this manner can differ substantially from those extracted from the last 10 years of a trend based on the entire record. All data reported in this appendix were obtained using WRTDS simulations based on the longest usable water-quality record available, as indicated in table 1. In cases where 10-year trends are needed, the trends should be determined using the last 10 years of the data tabulated in this appendix.
4. In the section “WRTDS Estimated Flux and Trend Results for Nutrients and Sediment for the Chesapeake Bay Nontidal Water-Quality Monitoring Network Through Water Year 2012,” flow-normalized trends were reported over the period-of-record for stations in the “*LONG_TERM*” scenario and over the most recent 10 years for stations in the “*10_YEAR*” scenario. Annual concentration and flux values, both raw and flow-normalized, are reported in this appendix for all stations, regardless of record length. However, because short records reflect a more limited range of hydrologic conditions, the USGS does not endorse reporting, or recommend interpreting, trends in flow-normalized concentration of flux for records shorter than 10 years in length.
5. Model results reported in this appendix do not reflect any adjustment to the period of reporting associated with known or suspected changes in storm sampling protocol, as discussed in the section “Variability in Annual Sampling Effort and Storm Sampling Effort.”

Appendixes (Continued)

Appendix 4. Effects of variability in annual sampling effort on trends—Graphic results for the eight watersheds subject to detailed study (available at <http://dx.doi.org/10.3133/sir20155133>)

Appendix 5. Effects of variability in storm sampling effort on trends—Graphic results for the eight watersheds subject to detailed study (available at <http://dx.doi.org/10.3133/sir20155133>)

Approved for publication on September 10, 2015

For more information about this publication contact:

Jeff Chanat
U.S. Geological Survey
Virginia Water Science Center
1730 East Parham Road
Richmond, VA 23228
jchanat@usgs.gov

Or visit the Virginia Water Science Center Web site
at <http://va.water.usgs.gov>

Prepared by the USGS Science Publishing Network,
Raleigh Publishing Service Center

I S B N 978-1-4113-4005-3



ISSN 2328-031X (print)
ISSN 2328-0328 (online)
<http://dx.doi.org/10.3133/sir20155133>