U.S. Department
of Transportation

**Federal Highway
Administration**

USGS
science for a changing world

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

Publication No. FHWA-HEP-09-003
2009

Methods for Development of Planning-Level Estimates of Water Quality at Unmonitored Stream Sites in the Conterminous United States

Office of Project Development
and Environmental Review
1200 New Jersey Avenue, SE
Washington, DC 20590

Foreword

The mission of the Federal Highway Administration (FHWA) is to continually improve the quality of our Nation's highway system and intermodal connections in a manner that protects and enhances the natural environment and communities affected by transportation. In enacting the Intermodal Surface Transportation Efficiency Act of 1991 (ISTEA); the Transportation Equity Act for the 21st Century (TEA-21) in 1998; and the Safe, Accountable, Flexible, Efficient Transportation Equity Act: A Legacy for Users (SAFETEA-LU) in 2005, the U.S. Congress has consistently emphasized the need for an integrated and multimodal transportation system that reflects environmental sensitivity and community values. Protecting and enhancing the environment and communities affected by transportation requires that principles of environmental stewardship be incorporated in all of the FHWA's policies, procedures, and decisions. This means that the FHWA responsibly considers and evaluates all aspects of the environment throughout the highway design, planning, and development process. Beyond its obligations embodied in environmental stewardship, the FHWA must demonstrate leadership on environmental matters in its collaboration with State and local agencies that implement transportation projects and programs throughout the country. The FHWA also has a responsibility to streamline the complex environmental stewardship process to ensure that highway projects are done in the most efficient and economical manner possible. To meet these goals, the FHWA must develop and disseminate research products that help FHWA and its partners implement surface transportation programs in a manner that protects and enhances the natural and human environment. More specifically, the Water and Ecosystems Team of the FHWA Office of Natural And Human Environment strives to develop and disseminate skills, tools, and information to redesign Federal environmental and transportation decisionmaking, and to ensure an integrated process at the Federal, State, tribal, and local levels. These tools, techniques and methods are designed to reduce direct and indirect adverse impacts of highways on water quality, habitat, and ecosystems to preserve and enhance human health, biological productivity, and ecological diversity.

This report, the associated computer applications, and data provide tools and techniques for developing planning-level estimates of background water-quality at sites receiving highway runoff. This information is vital for assessing the potential for adverse effects of runoff on receiving waters throughout the Nation. Ready availability of paired stream discharge and water-quality measurements in a standard format and the ease of use of the associated computer applications should provide transportation agencies with the tools and information to improve project delivery without compromising environmental protection. The methods described in this report, which focus on analysis of available water-quality data by ecoregion, are an embodiment of the Eco-logical approach for developing infrastructure projects (FHWA-HEP-06-011) developed by the FHWA in conjunction with the Council on Environmental Quality and more than 10 Federal and State agencies involved in transportation and natural-resource protection.

Patricia Cazenas
Highway Engineer, Water and Ecosystems Team
Office of Project Development and
Environmental Review
Federal Highway Administration

Carol Adkins,
Team Leader, Water and Ecosystems Team
Office of Project Development and
Environmental Review
Federal Highway Administration

Notice

This document is disseminated under the sponsorship of the U.S. Department of Transportation in the interest of information exchange. The U.S. Government assumes no liability for the use of the information contained in this document. This report does not constitute a standard, specification, or regulation.

The U.S. Government does not endorse products or manufacturers. Trademarks or manufacturers' names appear in this report only because they are considered essential to the objective of the document.

Quality Assurance Statement

The Federal Highway Administration (FHWA) provides high-quality information to serve Government, industry, and the public in a manner that promotes public understanding. Standards and policies are used to ensure and maximize the quality, objectivity, utility, and integrity of its information. FHWA periodically reviews quality issues and adjusts its programs and processes to ensure continuous quality improvement.

Technical Report Documentation Page

1. Report No. FHWA-HEP-09-003	2. Government Accession No.	3. Recipients Catalog No.
4. Title and Subtitle Methods for Development of Planning-Level Estimates of Water Quality at Unmonitored Stream Sites in the Conterminous United States		5. Report Date June 2009
		6. Performing Organization Code
7. Authors Gregory E. Granato, Carl S. Carlson, and Becca S. Sniderman		8. Performing Organization Report No.
9. Performing Organization Name and Address U.S. Geological Survey 12201 Sunrise Valley Drive Reston, VA 20192		10. Work Unit No. (TRAIS)
		11. Contract or Grant No. DTFH61-02-Y-30079
12. Sponsoring Agency Name and Address Federal Highway Administration Office of Project Development and Environmental Review 1200 New Jersey Avenue, SE Washington, DC 20590		13. Type of Report and Period Covered Final Report October 2003–September 2009
		14. Sponsoring Agency Code

15. Supplementary Notes

Contracting Officer's Technical Representatives (COTR's): Patricia A. Cazenas and Carol Adkins

16. Abstract

This report documents methods for data compilation and analysis of water-quality-transport curves that meet data-quality-objectives for order-of-magnitude planning-level estimates of stream-water quality at unmonitored sites in the 84 U.S. Environmental Protection Agency Level III nutrient ecoregions in the conterminous United States. The water-quality-transport curves developed in this analysis are intended for use with a stochastic data-generation algorithm, for use with a highway-runoff model designed to better quantify the risk of exceeding water-quality criteria as precipitation, discharge, ambient water quality, and highway-runoff quality vary from storm to storm. Transport curves are regression relations used to estimate constituent concentrations from measured or estimated water-discharge values. Three constituents, total phosphorus, total hardness, and suspended sediment, were selected for regression analysis to develop transport curves for each ecoregion. However, the data compilation and interpretation methods described herein may be used with other water-quality constituents. A total of 24,581 USGS surface-water-quality monitoring stations with drainage areas ranging from 0.002 to 1,140 square miles were identified in the conterminous United States and cataloged for retrieval of water-quality data. The number of paired water-discharge and water-quality samples for total phosphorus, total hardness, and suspended sediment concentrations was 246,403; 107,289; and 275,950, respectively. Examination of transport curves developed with these data indicate that these curves are appropriate models describing the underlying processes of washoff or dilution expected for each constituent, and that predictions made using these transport curves are comparable with published estimates for each water-quality constituent. All of the geographic information system files, computer programs, data files, and regression results developed for this study are included on the CD-ROM accompanying this report. The CD-ROM also contains a data directory with more than 1,876,000 paired discharge and water-quality measurements that include 21 other constituents commonly studied in highway- and urban-runoff studies.

17. Key Words transport curve, regression, ecoregion, event mean concentration, water quality, sediment quality, Stochastic Empirical Loading and Dilution Model	18. Distribution Statement No restrictions.

19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 53	22. Price

Form DOT F 1700.7 (8-72) (computer generated)

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS					APPROXIMATE CONVERSIONS FROM SI UNITS				
Symbol	**When You Know**	**Multiply By**	**To Find**	**Symbol**	**Symbol**	**When You Know**	**Multiply By**	**To Find**	**Symbol**
		LENGTH					**LENGTH**		
in	inches	25.4	millimeters	mm	mm	millimeters	0.039	inches	in
ft	feet	0.305	meters	m	m	meters	3.28	feet	ft
yd	yards	0.914	meters	m	m	meters	1.09	yards	yd
mi	miles	1.61	kilometers	km	km	kilometers	0.621	miles	mi
		AREA					**AREA**		
in^2	square inches	645.2	square millimeters	mm^2	mm^2	square millimeters	0.0016	square inches	in^2
ft^2	square feet	0.093	square meters	m^2	m^2	square meters	10.764	square feet	ft^2
yd^2	square yard	0.836	square meters	m^2	m^2	square meters	1.195	square yards	yd^2
ac	acres	0.405	hectares	ha	ha	hectares	2.47	acres	ac
mi^2	square miles	2.59	square kilometers	km^2	km^2	square kilometers	0.386	square miles	mi^2
		VOLUME					**VOLUME**		
fl oz	fluid ounces	29.57	milliliters	mL	mL	milliliters	0.034	fluid ounces	fl oz
gal	gallons	3.785	liters	L	L	liters	0.264	gallons	gal
ft^3	cubic feet	0.028	cubic meters	m^3	m^3	cubic meters	35.314	cubic feet	ft^3
yd^3	cubic yards	0.765	cubic meters	m^3	m^3	cubic meters	1.307	cubic yards	yd^3
	NOTE: volumes greater than 1000 L shall be shown in m^3								
		MASS					**MASS**		
oz	ounces	28.35	grams	g	g	grams	0.035	ounces	oz
lb	pounds	0.454	kilograms	kg	kg	kilograms	2.202	pounds	lb
T	short tons (2000 lb)	0.907	megagrams (or "metric ton")	Mg (or "t")	Mg (or "t")	megagrams (or "metric ton")	1.103	short tons (2000 lb)	T
		TEMPERATURE (exact degrees)					**TEMPERATURE (exact degrees)**		
^{o}F	Fahrenheit	5 (F-32)/9 or (F-32)/1.8	Celsius	^{o}C	^{o}C	Celsius	1.8C+32	Fahrenheit	^{o}F
		ILLUMINATION					**ILLUMINATION**		
fc	foot-candles	10.76	lux	lx	lx	lux	0.0929	foot-candles	fc
fl	foot-Lamberts	3.426	candela/m^2	cd/m^2	cd/m^2	candela/m^2	0.2919	foot-Lamberts	fl
		FORCE and PRESSURE or STRESS					**FORCE and PRESSURE or STRESS**		
lbf	poundforce	4.45	newtons	N	N	newtons	0.225	poundforce	lbf
lbf/in^2	poundforce per square inch	6.89	kilopascals	kPa	kPa	kilopascals	0.145	poundforce per square inch	lbf/in^2

*SI is the symbol for the International System of Units. Appropriate rounding should be made to comply with Section 4 of ASTM E380.

(Revised March 2003)

Contents

Figures

Tables

Conversion Factors and Water-Quality Units

Multiply	By	To obtain
	Length	
cubic foot per second (ft^3/s)	0.02832	cubic meter per second (m^3/s)
cubic foot per second per square mile($ft^3/s/mi^2$)	0.01093	cubic meter per second per square kilometer ($m^3/s/km^2$)

Water-Quality Units

Chemical concentrations in water are given in units of milligrams per liter (mg/L) or micrograms per liter (µg/L), which express the mass of solute per unit volume (liter) of water. Milligrams per liter are equivalent to “parts per million.” Micrograms per liter are equivalent to “parts per billion.” To calculate water-quality loads, there are 28.32 liters per second (L/s) in a cubic foot per second (ft^3/s) and 10.32 liters per second per square kilometer ($L/s/km^2$) in a cubic foot per second per square mile ($ft^3/s/mi^2$).

Acronyms

BCF	bias correction factor
BMP(s)	best management practice(s)
CCC	criteria continuous concentrations
CD-ROM	computer disk read only memory
CMC	criteria maximum concentrations
DQOs	data quality objectives
EMC(s)	event mean concentration(s)
FHWA	Federal Highway Administration
GIS	geographic information system
IQR	interquartile range
KTRLine	Kendall-Theil robust line
MAD	median absolute deviation
NASQAN	National Stream Quality Accounting Network
NWIS	National Water Information System
NWISSC	National Water Information System Site Cleaner
NWIS Web	National Water Information System Web
NWiz	National Water Information System Wizard
OLS	ordinary least squares
PDF	(Adobe) portable document format
PRESS	prediction error sum of squares
QA	quality assurance
RDBP	Relational DataBase File Processor
RMSE	root mean square error
RDB	relational database
SELDM	Stochastic Empirical Loading and Dilution Model
SQL	structured query language
STORET	STOrage and RETrieval database
SSC	suspended sediment concentration
SWQDM	Surface-Water-Quality Data Miner Database
TMDL(s)	Total maximum daily load(s)
TSS	total suspended solids
URL	uniform resource locator (Internet or Web address)
USEPA	U.S. Environmental Protection Agency
USGS	U.S. Geological Survey

Methods for Development of Planning-Level Estimates of Water Quality at Unmonitored Stream Sites in the Conterminous United States

By Gregory E. Granato, Carl S. Carlson, and Becca S. Sniderman

Abstract

This report documents methods for data compilation and analysis of water-quality-transport curves[1] that meet data-quality objectives for order-of-magnitude planning-level estimates of stream-water quality at unmonitored sites in 84 U.S. Environmental Protection Agency level III nutrient ecoregions in the conterminous United States. Water-quality- transport curves, which are regression relations used to estimate constituent concentrations from measured or estimated water discharge values, are intended for use with a stochastic data-generation algorithm. These transport curves can be used with a highway-runoff model to better quantify the risk of adverse effects on receiving-water quality, which varies from storm to storm with variations in precipitation, discharge, ambient water quality, and highway-runoff quality.

Water-quality-transport curves were developed using the nonparametric Kendall-Theil robust line method as implemented in a software program (KTRLine version 1.0) developed for this study. Transport curves were developed because concentrations of many constituents commonly vary as a result of washoff and dilution processes in receiving waters. Three constituents—total phosphorus, total hardness, and suspended sediment—were selected for regression analysis to develop transport curves for each ecoregion; however, these data compilation and interpretation methods may be used with other water-quality constituents. These methods also may be used with other information, such as local land-use data, for more selective regional or site-specific analysis. All the geographic-information-system files, computer programs, data files, and regression results developed for this study are included on the CD-ROM accompanying this report. The CD-ROM also contains a data directory with more than 1,876,000 paired discharge and water-quality measurements that include 21 constituents commonly studied in highway- and urban-runoff studies.

This national synthesis effort was based on data available on the U.S. Geological Survey (USGS) National Water Information System Web. A total of 24,581 USGS surface-water-quality monitoring stations with drainage areas ranging from 0.002 to 1,140 square miles were identified in the conterminous United States and cataloged for retrieval of water-quality data. The number of these stations with paired water-discharge (measured streamflow per unit time) and concentration data is: total phosphorus (parameter code p00665)—8,169; total hardness (parameter code p00900)—7,290; and suspended sediment concentrations (parameter code p80154)—7,477. The number of paired water-discharge and water-quality samples is: total phosphorus—246,403; total hardness—107,289; and suspended sediment concentrations—275,950. Most (98 percent) of the instantaneous water discharge values in the data set ranged from 0.001 to 50 cubic feet per second per square mile. Most (98 percent) of the concentration values in the data set ranged from 0.005 to 5 milligrams per liter (mg/L) for total phosphorus; 4 to 3,000 mg/L for total hardness, and 0.6 to 30,000 mg/L for suspended sediment concentrations. About 9 percent of total phosphorus values, about 0.1 percent of total hardness values, and about 1 percent of suspended sediment concentration values are identified at less than detection limits, greater than maximum reporting limits, estimated, or some other qualification code. Examination of transport curves developed with these data indicate that these curves are appropriate models describing the underlying processes of washoff or dilution expected for each constituent, and that estimates of concentrations made using these transport curves are comparable to published estimates for each water-quality constituent.

---

[1] Many technical terms are listed in the glossary at the back of this report.

Introduction

Planning-level estimates of event mean concentrations (EMCs) of water-quality constituents in runoff and receiving waters at unmonitored sites can be used to evaluate potential effects of highway and urban runoff in receiving waters (Marsalek, 1991). EMCs are operationally defined as the total water-quality constituent load from a storm event divided by the total volume of runoff from a storm. EMCs are commonly estimated from flow-proportional water-quality sampling programs. Planning-level estimates are commonly defined as the results of analyses used to evaluate broad policy measures. Planning-level estimates are recognized to include substantial uncertainties (commonly orders of magnitude) in all aspects of the decision process (Barnwell and Krenkel, 1982; Marsalek and Ng, 1989; Marsalek, 1991). For example, in a review of different stormwater modeling methods, Chandler (1994) determined that differences between results of selected modeling methods were less than the variability in stormwater EMCs, which commonly varied by orders of magnitude at each site.

Estimates of the frequency and magnitude of potential adverse effects on receiving-water quality are used in planning efforts to assess potential needs for various mitigation strategies, such as the implementation of structural best management practices (BMPs), as part of a design solution where water-quality problems have been identified. Water-quality criteria are used to estimate the potential for adverse effects of selected constituents in receiving waters. Water-quality criteria are expressed as criteria maximum concentrations (CMC) and criteria continuous concentrations (CCC) of constituents of concern in receiving waters downstream from point- or nonpoint-source discharges. The CMC are associated with acute toxicity, which is defined as constituent concentrations causing death from short-term (a period of hours) exposures. The CCC are associated with chronic toxicity, which is defined as constituent concentrations causing lethal and nonlethal effects (such as reduced growth or reproduction) from exposure over longer periods of time (days or weeks). These criteria are commonly applied as a function of the duration and recurrence of concentrations that exceed the criteria in the receiving water (usually once in 3 years). Decisionmakers use a specified estimate of water discharge (measured streamflow per unit time) and upstream water-quality constituent concentrations to estimate allowable concentrations and flows for discharges to the receiving waters (U.S. Environmental Protection Agency, 1986b; U.S. Environmental Protection Agency, 2002). The U.S. Department of Transportation, Federal Highway Administration (FHWA), however, commonly plans mitigation strategies to minimize the potential for adverse effects from highway runoff, even if criteria exceedences are improbable (P.A. Cazenas, Federal Highway Administration, written commun., 2006).

A mass-balance approach (fig. 1) is commonly applied to estimate the concentrations and loads of water-quality constituents in receiving waters downstream of an urban or highway-runoff outfall (Warn and Brew, 1980; Di Toro, 1984; Driscoll and others 1989; Driscoll and others 1990c). In a mass-balance model, the loads (the product of measured water discharge times concentration) of the upstream and runoff discharge components are summed to calculate the discharge, concentration, and load in the (fully mixed) receiving water downstream of a discharge point. Application of a mass-balance model, however, requires that statistics describing the frequency distributions of component discharges and concentrations be used to determine the statistics for downstream discharges, concentrations, and loads (Warn and Brew, 1980). For example, DiToro (1984) used information about probability distributions of EMCs and storm-runoff discharges and upstream receiving waters to develop an empirical probabilistic-dilution model that was used to develop planning-level estimates of downstream EMCs and water discharges. The resulting probability distribution of downstream EMCs indicates the potential for exceeding water-quality criteria and, therefore, the potential need for more information and data that may be used to identify suitable mitigation measures. DiToro (1984) based his method on the assumption that contributing discharges and concentrations are independent and lognormally distributed. Warn and Brew (1980), however, indicated that upstream concentrations and loads are correlated.

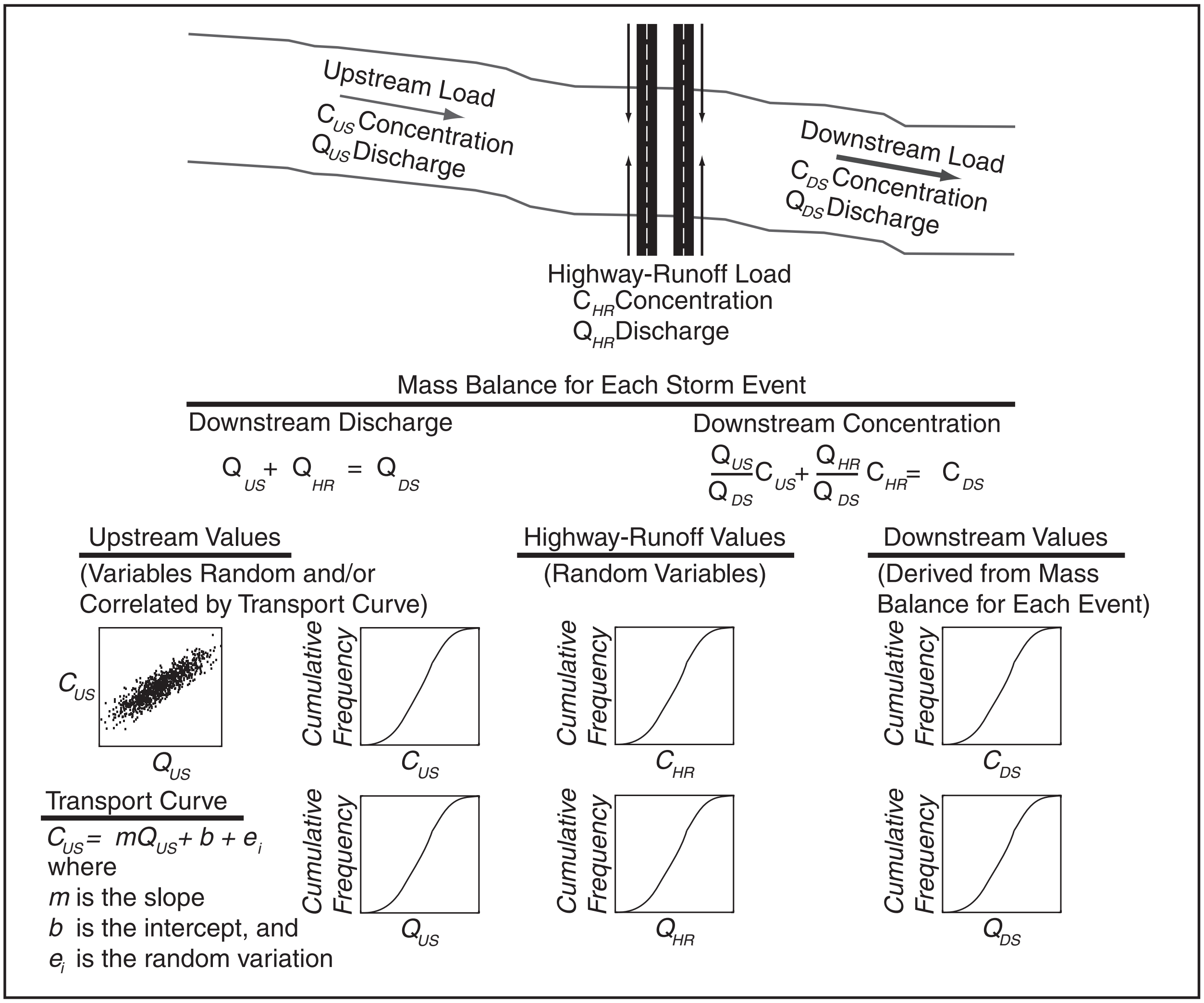


Figure 1. Schematic diagram showing the stochastic mass-balance approach for estimating discharge, concentration, and loads of water-quality constituents upstream of a highway-runoff outfall, from the highway, and downstream of a highway-runoff outfall.

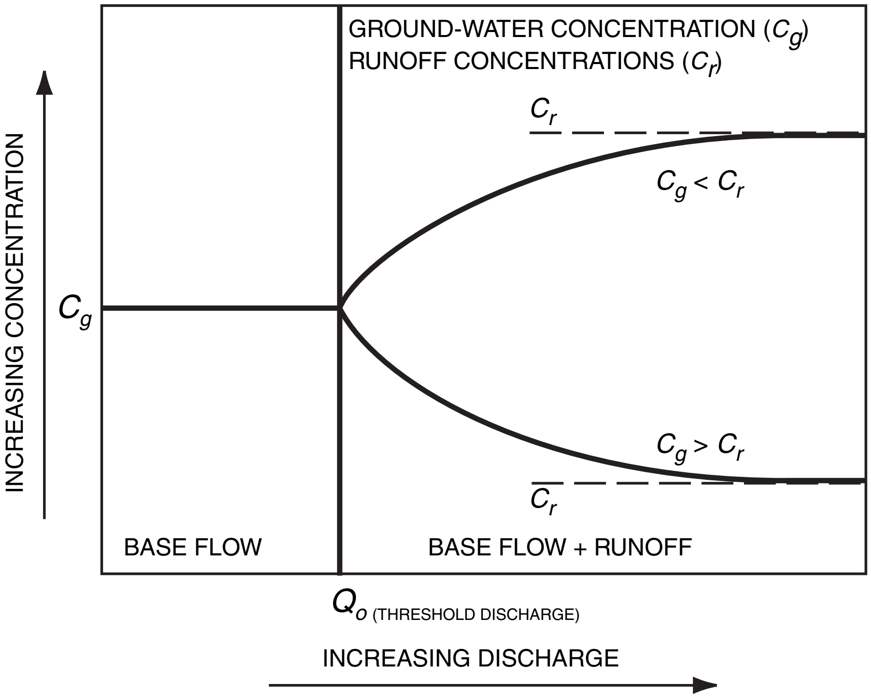


Figure 2. Schematic diagram showing simplified relations between increasing discharge and the concentration of constituents of interest in a stream or river during base flow and runoff events. (Modified from O'Connor, 1976.)

Concentrations of sediment, sediment-associated constituents (selected nutrients, organic compounds, and trace elements), and dissolved constituents in receiving waters commonly vary as result of washoff and dilution processes. Relations between concentration and discharge commonly are referred to as water-quality-transport curves and are an accepted method for characterizing water-quality variation with discharge (Glysson, 1987). Constituent concentrations commonly vary with water discharge because these constituents either are mobilized from the land surface during storms (washoff) or are present in receiving waters from other sources and are diluted by less concentrated storm runoff (dilution). These processes produce variations in constituent concentrations that are a function of water discharge. The transport curve shown schematically in figure 2 is an example of a washoff constituent such as sediment that commonly increases in concentration with increasing discharge.

O'Connor (1976) set forth a simplified conceptual model (fig. 2) to illustrate these processes. In this model, the stream has a relatively constant concentration (C_g) below a threshold water discharge (Q_o), which is commonly referred to as base flow. If washoff is the dominant process controlling a constituent concentration, the concentration of the constituent in surface runoff (C_r) is greater than that in base flow. In this case, solutes, sediment, and sediment-associated constituents are delivered to the stream by overland flow or by erosion of bed or bank materials. For constituents dominated by washoff or erosion, concentration and water discharge are positively related and have a positive slope with increasing water discharge above Q_o. If dilution is the dominant process controlling constituent concentrations, the concentration of the constituent in base flow is greater than the concentration in surface runoff, and the relation between concentration and water discharge will have a negative slope with increasing water discharge above Q_o. In either case, concentrations of constituents are expected to approach concentrations in runoff as the volume of runoff increases (fig. 2). Although base-flow concentrations (C_g) and threshold flows (Q_o) vary seasonally, and many other factors complicate the analysis, this type of multisegment model has been shown to be useful for dissolved constituents (O'Connor, 1976), suspended sediment (Glysson, 1987; Simon, 1989), and sediment-associated constituents (Smith and others, 1982). Some constituents at some sites show combinations of both behaviors, and therefore, relations can have segments of downward and upward slope (Hirsch and others, 1991).

Regression equations, referred to as water-quality-transport curves, may be used to quantify correlations between upstream discharges and concentrations (O'Connor, 1976; Glysson, 1987). A water-quality-transport curve, however, will provide only one unique value of concentration for each discharge value rather than the random distribution of concentrations above and below the transport curve that is characteristic of water-quality data. If it is assumed that regression residuals are randomly distributed above and below the water-quality-transport curve, regression statistics may be used with Monte Carlo methods for stochastic data generation to model the population of upstream concentrations and water discharges as a transport curve with a random-error component (fig. 1). These estimates of upstream concentration and discharge may be used with statistics describing the population of highway-runoff discharges and constituent concentrations (Granato and Cazenas, 2006) to derive mass-balance estimates of the population of downstream concentrations and discharges.

In 1990, the FHWA published a highway-runoff water-quality model (Driscoll and others, 1990a, b, c, d). This water-quality model is useful for developing planning-level estimates of selected highway-runoff constituent concentrations and estimating the potential effectiveness of selected BMPs. The 1990 model, however, was a dilution model based on the assumption that concentrations of the constituents of concern in the upstream receiving waters could be approximated as zero. A new model that could incorporate nonzero upstream concentrations and discharges with estimates of highway concentrations and discharges would improve the planning-level estimates of the resultant concentrations and discharges in downstream receiving waters. Information about receiving-water quality and water discharge is necessary to estimate the probability distributions of constituent concentrations and loads upstream of a highway-runoff outfall to implement such a model. For example, Driscoll and others (1989) implemented DiToro's (1984) methods in a stochastic probability dilution model (PDM) for the U.S. Environmental Protection Agency (USEPA). The PDM model, however, is not widely cited in the literature and was not included in a recent compendium of modeling methods for watershed assessment and total maximum daily load (TMDL) development published by the USEPA (Shoemaker and others, 1997). Use of the PDM model for planning-level estimates of receiving-water quality has been limited by the expense and difficulty of obtaining suitable data for estimating upstream concentrations in receiving waters (E.D. Driscoll, oral commun., 2002).

The U.S. Geological Survey (USGS), in cooperation with the FHWA, began a study to develop the stochastic empirical loading and dilution model (SELDM) in 2003. SELDM is a water-quality model that utilizes available data and stochastic Monte Carlo methods to generate planning-level estimates of water-quality constituent concentrations, discharges and loads from the watershed upstream of the highway, from the highway itself, and in the receiving water downstream from the highway discharge for storm events. This information can be used to evaluate highway runoff as a source of water-quality constituents, the potential effects of these loads on receiving-water quality, and the potential effectiveness of BMPs to control the highway-runoff portion of the downstream water discharge, constituent concentrations, and loads. A runoff-quality model based on nonzero upstream concentrations creates a need for national, regional, and local estimates for these upstream concentrations. The effort, time, and expense to collect and analyze concentration data, however, limit the availability of water-quality measurements that may represent conditions for a given site. Therefore, methods to develop robust planning-level estimates of the constituent concentrations in receiving waters as a function of water discharge at unmonitored sites across the country would be needed. These estimates would be made on a regional basis with data that are available from a reliable source for a national-scale model. Runoff-quality models can be used to estimate constituent concentrations upstream and downstream from a highway-runoff outfall, if such data are available.

Purpose and Scope

This report documents the compilation and analysis of a national data set of paired values of instantaneous water discharge and concentrations of selected water-quality constituents. Data from across the conterminous United States were grouped by 84 USEPA level III nutrient ecoregions (U.S. Environmental Protection Agency, 2003b, 2004). Surface-water-quality sampling sites were identified to help characterize water quality for regional or local analysis. Water-quality-transport curves developed in this analysis were intended for use with a stochastic data-generation algorithm. The data-generation algorithm was designed to provide planning-level estimates of water-quality concentrations from estimates of water discharge. These estimates may be used in a Monte Carlo analysis (Haan, 1994) of water quality immediately upstream from a site where runoff from a highway may drain to a receiving water body. This compilation and analysis effort included data from almost all available water-quality-monitoring stations in an attempt to better represent ambient water-quality-conditions immediately upstream of any runoff outfall. The methods described in this report may be used with a subset of water-quality-monitoring stations that are minimally affected by anthropogenic activities (such as the sites used by Smith and others, 2003) to represent natural background water quality. These methods also may be used with selected local stations or data collected at the site of interest to refine estimates of ambient water quality.

The current study provides methods to derive order-of-magnitude planning-level estimates of EMCs of ambient receiving-water quality for unmonitored sites in the conterminous United States. This study also provides methods that may be useful to obtain and interpret more quantitative site-specific data. If the regional water-quality estimates described in this report do not meet data-quality objectives (DQOs) for a particular project, users may refine estimates by selecting and analyzing site-specific data. Existing site-specific data may be obtained from the USGS on-line interface to the National Water Information System (NWIS Web) (U.S. Geological Survey, 2004a). Alternatively, if DQOs for a particular project require site-specific data, the users may obtain their own data and use the tools and methods described in this report to record and analyze that data to refine estimates of ambient water quality at any given site. Regional estimates may be more robust for predicting environmental variables at unmonitored sites than measurements from a relatively short-duration site-specific sampling program, unless the sampling program characterizes the full range of water discharge and is not affected by short-term natural or anthropogenic factors (Hughes and Larsen, 1988; Hosking and Wallis, 1997; Vogel and others 1998; Robertson and others, 2001; Shirazi and others, 2001; Jenerette and others, 2002).

Four water-quality constituents, total nitrogen (parameter code p00600), total phosphorus (parameter code p00665), total hardness (parameter code p00900), and suspended sediment concentrations (parameter code p80154) were selected for analysis on the basis of their perceived importance for current water-quality analysis. Indicators of the perceived importance of a constituent included use of the constituent in the 1990 FHWA water-quality model (Driscoll and others, 1990c), and the international stormwater BMP database (Urbonas, 1996). Inclusion in the list of total maximum daily loads (TMDLs) also was a criterion for selection (U.S. Environmental Protection Agency, 2006). The final criterion for constituent selection was potential availability of data sets from water-quality-monitoring stations in all 84 level III nutrient ecoregions of the conterminous United States. Three water-quality constituents, total phosphorus (parameter code p00665), total hardness (parameter code p00900), and suspended sediment concentrations (parameter code p80154), were used for development of regional water-quality-transport curves because of the availability of a substantial amount of data in all 84 level III nutrient ecoregions of the conterminous United States. More than 1,876,000 paired discharge and water-quality measurements that include 21 other constituents commonly measured in highway- and urban-runoff studies also were retrieved, processed, and saved on the CD-ROM accompanying this report.

Several computer programs and database applications were developed, used, and documented for this analysis and for future use. These computer programs and database applications include:

- National Water Information System Wizard (NWiz version 1.0)–A program used to retrieve and examine information and data from NWIS Web;
- National Water Information System Site Cleaner (NWISSC version 1.0)–A program used to clean up files that do not include required data from directories containing the results of water-quality data queries from NWIS Web;
- Relational DataBase File Processor (RDBP version 1.0)–A program used to select and reformat water-quality data from NWIS Web; and
- Surface-Water-Quality Data Miner Database (SWQDM version 1.0)–A database application to facilitate storage and use of surface-water-quality and discharge data for national, regional, or local analysis.

These computer programs and database applications, which are designed to be used together to obtain, process, and catalog data, are documented on the CD-ROM accompanying this report along with the results of the data selection, compilation, and interpretation effort.

Data Availability

The USGS National Water Information System (NWIS) internet application, NWIS Web, is a source of water-quality data that can be used to estimate local, regional, or national water-quality parameters (Mathey, 1998; U.S. Geological Survey, 2002; 2004a). NWIS Web includes millions of water-quality records from hundreds of thousands of sites in all 50 states (Turcios and Gray, 2001). Review of the data, however, indicates that many sites in the database have relatively few water-quality samples for any given constituent. For example, in a review of sediment data in the NWIS Web database, Turcios and Gray (2001) determined that less than 25 percent of monitoring sites with paired suspended sediment and water discharge measurements had 30 or more available measurements. Therefore, regionalization or combining data from nearby sites may be necessary to produce quantitative estimates of water quality at unmonitored sites from available data sets.

Data also may be available from the U.S. Environmental Protection Agency (USEPA) STOrage and RETrieval (STORET) database (U.S. Environmental Protection Agency, 2005a). The STORET Web site includes legacy data collected from the 1960s through 1999 and an option for a modernized STORET system with data collected since 1999. As of 2006, the STORET database does not include data for large parts of the U.S., including parts of Alabama, California, Idaho, Illinois, Indiana, Massachusetts, Mississippi, Nevada, New Mexico, New York, Oregon, Texas, Virginia, and Washington. The STORET database also includes data from wells, wastewater-treatment plants, USEPA Superfund sites, landfills, and mine discharge, which are important on a local scale but may skew estimates of ambient receiving-water quality for a regional analysis. The USGS NWIS and the USEPA STORET databases use common definitions and formats to promote a common view of data between the two systems. The methods described in this report may be adapted for selective use with STORET data.

Data-Quality Objectives

The DQO process is designed to help evaluate the costs of data acquisition in relation to the consequences of a decision error caused by inadequate input data (U.S. Environmental Protection Agency, 1986a; 1994; 1996; Granato and others, 2003). DQOs are meant to ensure that data and interpretations are useful for the intended purpose. In a review of water-quality data collected by federal, state, and local water-quality monitoring entities, Hren and others (1987) defined five characteristics necessary to establish that data are useful. To be useful, data must be (1) representative of the system under study; (2) available for public use as original data; (3) collected from a readily located sampling site (to assess data comparability and to interpret results of geographic/climatological variations); (4) associated with sufficient quality-assurance (QA) and quality control (QC) information (to indicate the validity, reliability, and compatibility of data from different sources); and (5) available in useful computer files (to increase reliable compilation and manipulation of large volumes of data). The methods and data developed and described in this report to obtain and interpret paired values of instantaneous water discharge and selected water-quality constituent concentrations meet these five characteristics because the data are obtained from the USGS NWIS Web database. This database includes information about the water-quality constituent of interest, the sample matrix, the data-collection entity, as well as the location, date, and time of sample collection. Because values in NWIS Web have been approved for public release, it is assumed that the data have been collected, handled, and analyzed with documented standard methods (Mathey, 1998; U.S. Geological Survey, 2002).

The FHWA has established a system of water-quality assessment and action plans that include different levels of interpretive analysis to determine potential environmental effects of highway runoff (Sevin, 1987; Cazenas and others, 1996; Federal Highway Administration, 1998). DQOs for these assessments depend on the level of interpretive analysis, which ranges from a completely qualitative initial assessment through an increasingly quantitative series of planning-level estimates. These planning-level estimates may be based on values derived from the literature, regional statistics, existing data for a given area, and site-specific data. The FHWA water-quality-assessment process is a step-by-step decision tree (Sevin, 1987; Cazenas and others, 1996; Federal Highway Administration, 1998). In the FHWA process, an initial assessment is completed to estimate the probability that the highway configuration being considered will produce unacceptable environmental effects. If the risk of an adverse effect is unacceptable to decisionmakers, the assessment is refined with more detailed data and analysis.

The process is concluded if it can be demonstrated that there is a low probability that implementation of the highway design (including proposed BMPs) would produce unacceptable environmental effects (Sevin, 1987; Cazenas and others, 1996; Federal Highway Administration, 1998). The decision rule for DQOs in this process is dependent on the sensitivity of the receiving waters, the presence of water supplies in the watershed, uncertainties associated with available data, and limitations of the analysis (P.A. Cazenas, Federal Highway Administration, oral commun., 2005). This compilation and interpretation of national water-quality data sets is designed to meet DQOs for development of planning-level estimates of stream-water quality at unmonitored sites in the conterminous United States.

Previous Investigations

Estimates of concentrations and loads of sediment and other water-quality constituents from water discharge measurements is of particular interest because continuous records of daily water discharge are available at more than 20,000 USGS streamflow-gaging stations in the United States for periods of years to decades, but records of measured constituent concentrations are relatively sparse. The ability to estimate concentrations and loads of water-quality constituents as a function of water discharge is considered important in studies to assess effectiveness of programs for abating nonpoint-source pollutants and to understand the transport and fate of sediment-borne constituents (Crawford, 1991). Because water discharge also can be estimated at ungaged sites, regional-transport curves may be used to estimate water quality at such sites. Regression models of water quality and water discharge also are used for data generation in the development of stochastic models (Koch and Smillie, 1986). Many studies have examined regression methods for the analysis of water-quality variables such as sediment (Miller, 1951; Glysson, 1987; Clarke, 1990a, b; Gilroy and others, 1990; Crawford, 1991; Nash, 1994; Syvitski and others, 2000; Vogel and others 2003, 2005), nutrients (Smith and others, 1982; Clarke, 1990a; Cohn and others, 1992; Cohn, 1995; House and Warwick, 1998), and major ions (O'Connor, 1976; House and Warwick, 1998; Albek, 1999, 2003). Many natural and anthropogenic influences, however, may affect relations between water discharge and water quality. Natural processes may include seasonality (Glysson, 1987; Cohn and others, 1992) and hysteresis in concentrations in the rising and falling parts of the hydrograph (O'Connor, 1976; Glysson, 1987; House and Warwick, 1998). Natural and anthropogenic influences include changes in watershed characteristics with time (Glysson, 1987; Simon, 1989; Cohn and others, 1992).

Methods of Data Selection, Compilation, and Interpretation

Standard methods are used for selecting, compiling, and interpreting available water-quality and water-discharge data from surface-water-monitoring stations. These methods must meet DQOs for derivation of regional planning-level estimates of receiving-water quality at sites throughout the conterminous United States. This report presents a standard automated process designed to facilitate regional or national synthesis effort by use of geographic-information system (GIS) files and several computer applications (fig. 3). This process includes methods for regionalization and site selection, data compilation, and data interpretation. After plausible regions are defined, sufficient numbers of water-quality-monitoring sites with data that can be used to indicate regional water-quality characteristics are identified. Data compilation is done to obtain, process, and format the required data. Data interpretation is done to select data by region, apply a nonparametric regression-analysis technique to the paired water-quality and water-discharge data, and to estimate regional water-quality characteristics by use of the regression equations.

Regionalization and Site Selection

Regional estimates of ambient surface-water quality are used to characterize receiving waters and the potential effect of nonpoint-source runoff on the ecology of these waters at unmonitored sites (Tasker and Granato, 2003). Methods for selecting regions include detailed statistical analysis of the parameter of interest (Langbein and Schumm, 1958; Hosking and Wallis, 1997; Lins, 1997; Robertson and others, 2001; Jenerette and others, 2002), quantitative regionalization of explanatory variables for the parameter of interest (Robertson and others, 2001; Griffith and others, 2002; Smith and others, 2003a), semi-quantitative regionalization of explanatory variables for the parameter of interest (Hughes and Larsen, 1988; Omernik, 1995; Omernik and Bailey, 1997; Omernik and others, 2000; Rohm and others, 2002), contour maps of average concentrations (Rainwater, 1962; Omernik and Powers, 1983; U.S. National Atmospheric Deposition Program, 2004), and physical, political, or administrative boundaries, such as states or hydrologic units (U.S. Geological Survey, 1982; Omernik and Bailey, 1997). Each method for regionalizing environmental data has benefits and limitations, but the objective of each is to reduce variability in relations between discharge measurements and constituent concentrations in a national data set by identifying areas with common characteristics that may influence these relations.

The U.S. Environmental Protection Agency (2003b; 2004) level III nutrient ecoregions were selected to regionalize data for this study because they provide a consistent national context for regionalization of environmental data acceptable

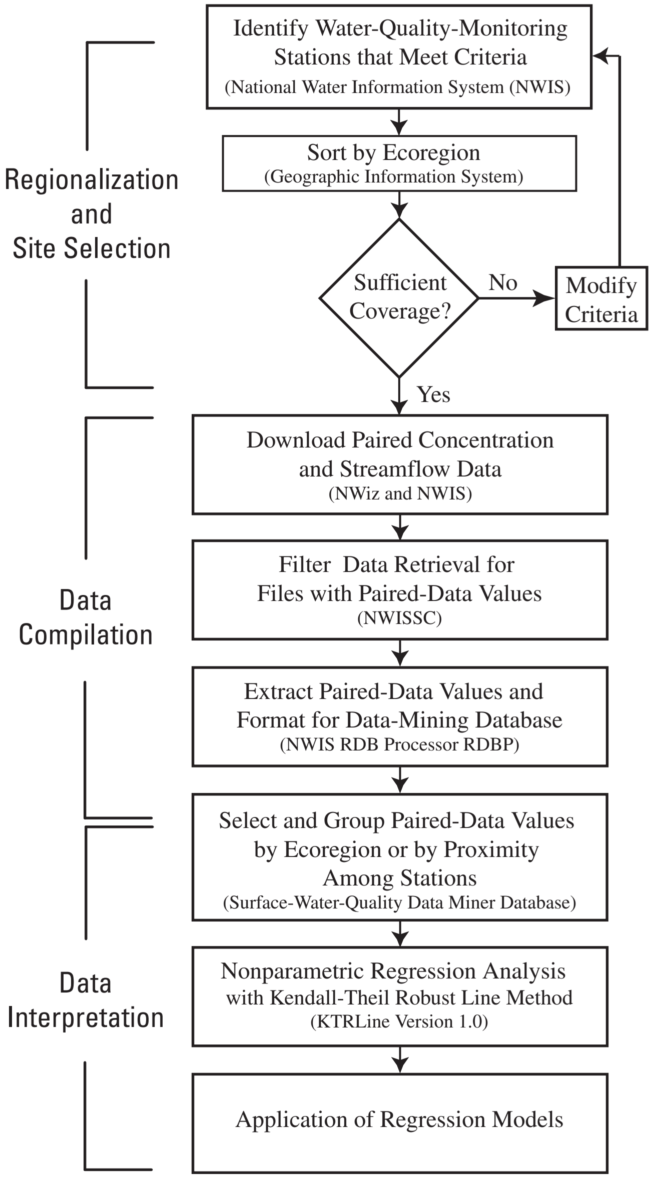


Figure 3. Process-flow diagram of the steps (and associated software) used for regionalization and site selection, data compilation, and data interpretation to define relations between surface-water discharge and concentrations of selected water-quality constituents for use in a national runoff-quality model.

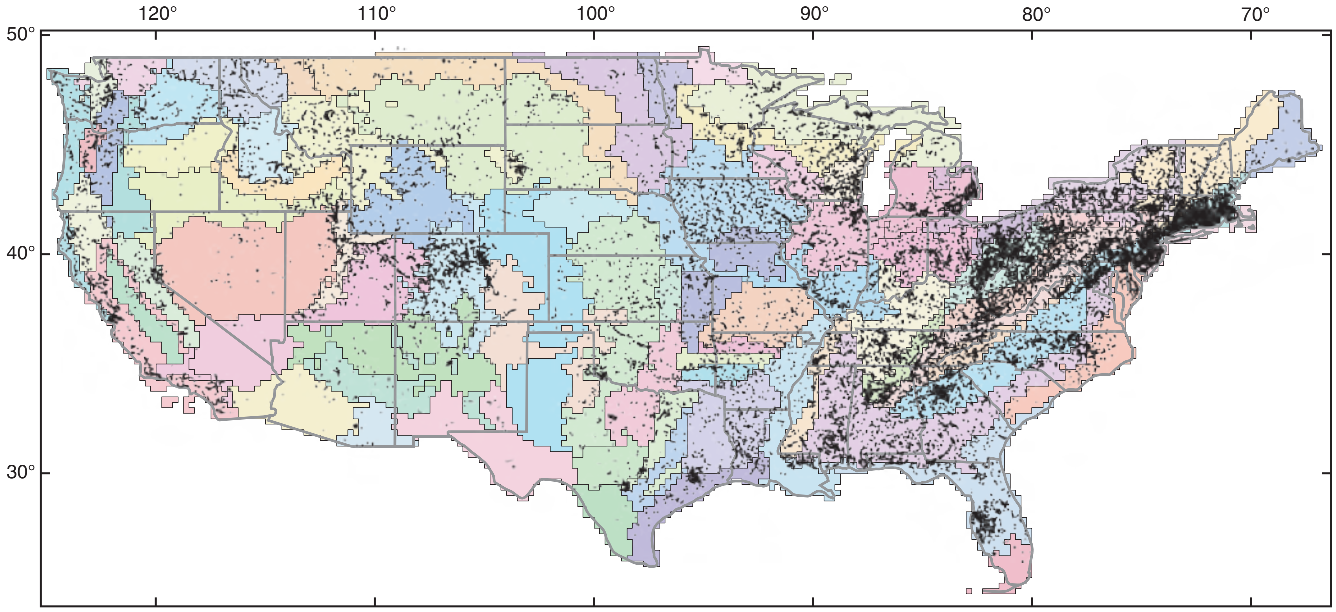


Figure 4. Spatial distribution of 24,581 surface-water-quality-monitoring stations (black dots) with drainage areas less than 1,050 square miles among U.S. Environmental Protection Agency (2003b) Level III ecoregions (colored polygons) that have been discretized to a 15-minute grid in the conterminous United States (geographic projection).

to decisionmakers and regulators for developing estimates of environmental conditions within each region for planning-level runoff-quality analysis. Ecoregions are defined as areas of relative homogeneity in ecological systems and their components (Omernik and others, 2000; Omernik, 2004). Environmental-resource-management agencies in many states increasingly are using ecoregions to set water-quality criteria, develop biological criteria, and evaluate nonpoint-source management goals (Omernik and Bailey, 1997). Federal agencies that have responsibilities for water-quality monitoring or management also are using ecoregions as a spatial framework to organize and interpret environmental data (Intergovernmental Task Force on Monitoring Water Quality, 1995a, b). For example, the FHWA in conjunction with the Council on Environmental Quality and more than 10 federal and state agencies involved in transportation and natural-resource protection has developed the “Eco-logical” approach for developing infrastructure projects that is based on analysis by ecoregion (Brown, 2006).

Delineation of ecoregions is a semi-quantitative or qualitative process that uses information about geology, physiography, vegetation, climate, soils, land use, wildlife, and hydrology (Omernik, 1995; Omernik and others, 2000). Historically, these factors have been used as explanatory variables in models for predicting water discharge (Thomas and Benson, 1970; Jennings and others, 1994; Vogel and others, 1998) and water quality (Langbein and Schumm, 1958; Syvitski and others, 2000; Shirazi and others, 2001; Griffith and others, 2002; Smith and others, 2003a). The ecoregion classification system has been demonstrated to be a useful indicator of regional water-quality characteristics (Heiskary and others, 1987; Hughes and Larsen, 1988; Robertson and others, 2001; Rohm and others, 2002; Simon 2003; Simon and others, 2004). Use of site-specific explanatory variables, such as local land use or soil types, commonly improves generalized predictions based on ecoregion-scale characteristics (Jenerette and others, 2002; Smith and others, 2003a). Detailed site-specific information, however, is not currently available in a uniform national data set for the USGS data during the period of record of available water-quality data.

To facilitate database analysis of ecoregion data, the USEPA ecoregion coverage was discretized to the resolution of 15-minute latitude-longitude grid squares during the current investigation (fig. 4). The region that had the highest proportion of area in each grid square was assigned as the ecoregion for the grid square. This discretization process does not violate the intent of the original delineation, because the 0.25-decimal-degree (15-minute) grid square is relatively small in comparison to the scale of the ecoregions. Although ecoregion boundaries are shown on maps as lines, transitions between regions actually occur over large amorphous zones along the edges of each region (J.M. Omernik, U.S. Environmental Protection Agency, written commun., 2004). Information about the USEPA (2004) level III nutrient ecoregions, including a printable ecoregion map, GIS

Table 1. Geographic-information system (GIS) files used for regionalization and site selection.

[USEPA, U.S. Environmental Protection Agency; USGS, U.S. Geological Survey; --, not applicable]

Data layer name	Scale	Feature type	Citation	Short description
gridtemplate	1:250,000	polygon	--	Fishnet grid of 0.25-decimal-degree (15-minute) latitude-longitude grid cells.
ecoregions	1:250,000	polygon	USEPA, 2003	USEPA level III nutrient ecoregions discretized to the 0.25-decimal degree grid.
waterquality	1:250,000	point	USGS, 2004	Water-quality-monitoring station information.

coverages (table 1) of the discretized version of the level III nutrient ecoregions in the conterminous United States, and a copy of the discretization grid, is provided on the CD-ROM accompanying this report.

Sites were selected by choosing surface-water-quality-monitoring stations that have a defined drainage area and at least one water-quality measurement with an associated water-discharge measurement defined in NWIS. A total of 24,581 water-quality-monitoring stations throughout the conterminous United States were identified and cataloged for retrieval of water-quality data from NWIS Web (fig. 4). A GIS coverage that identifies station location is provided on the CD-ROM accompanying this report (table 1). NWIS Web includes basic information about water-quality-monitoring stations, such as the location, altitude, drainage area, and topographic setting (U.S. Geological Survey, 2004a). NWIS Web does not currently include information about other explanatory variables that are used for ecoregion analysis, such as local soils or land use. The data on the CD-ROM, however, can be used with local GIS coverages that include such data. Station location coordinates can be used to assign each station to the appropriate ecoregion. The focus of much of the USEPA analysis efforts has been to develop reference or “minimally impacted” conditions by selecting reference sites or by choosing the lowest quartile of data from all sites in a region (U.S. Environmental Protection Agency, 2000; Robertson and others, 2001). Studies indicate that most available water-quality data sets are collected for a specific purpose, usually in an area where there is a perceived problem (Hren and others, 1987; Norris and others, 1990). Relatively few sites are monitored to determine reference conditions. For example, Smith and others (2003a) extrapolated estimates of natural background concentrations of nutrients in streams and rivers of the conterminous United States from 63 stations that meet all their criteria for “minimally impacted” reference basins.

The investigation described in this report was designed to provide generalized estimates of ambient receiving-water quality by ecoregion without incorporation of site-specific characteristics. Use of reference basins that represent natural background concentrations, however, may lead to underestimation of water-quality constituent concentrations commonly found in streams receiving highway runoff. Thus, data from reference basins may, on a regional basis, underrepresent the potential for adverse water-quality effects from highway runoff. Conversely, selection of data from sites that are substantially affected by anthropogenic activities may, on a regional basis, overrepresent the potential for adverse water-quality effects from highway runoff. Use of all available surface-water-quality data is warranted to best represent potential constituent concentrations in receiving waters. Consequently, the first objective of the site-selection process was to provide a sufficient number of water-quality-monitoring stations in each ecoregion to provide adequate data to develop relations between individual measurements of water discharge and water-quality data. The site-selection process was designed to be inclusive so that the resultant data set could be used to estimate water quality in any basin that may be affected by highway runoff.

The second objective of the site-selection process was to minimize the drainage-basin area of sites in the data set. Larger basins are more likely to include impoundments that alter the flow regime and affect water quality (Meade and Parker, 1985; Smith and others, 2003). Smaller basins are more likely to contain fewer variations in the explanatory variables that affect water quality. Use of smaller basins minimizes the effects of in-stream physical and chemical processes that cause variations in water quality (Smith and others, 2003a, b). Conceptually, it also is desirable to have data that represent conditions in small basins, where highway runoff may have a measurable effect on downstream water quality (Driscoll and others, 1990c).

In practice, balancing the need for data in each ecoregion with the desire to minimize drainage-basin area made site selection an iterative process (fig. 3). NWIS Web was queried to retrieve all surface-water-quality-monitoring stations with less than a specified maximum drainage area. After each NWIS query, a GIS query of the updated data set was run to count the number of monitoring sites in each ecoregion. During this iterative process, the specified maximum drainage-basin size was increased from an initial value of 50 square miles (mi^2) to 1,140 mi^2 in an effort to identify multiple monitoring stations with sufficient water-quality data in each level III ecoregion.

Data Compilation

Methods and tools for automated compilation of water-quality data were developed to establish a preliminary database of selected water-quality constituents and to provide the means for future data-compilation efforts that may focus on a particular site or a particular water-quality issue. The NWIS Web database (U.S. Geological Survey, 2004a) contains the data necessary to characterize measured water quality at many sites and to estimate water quality at similar sites that do not have a record of water-quality data. The interactive NWIS Web interface, however, does not facilitate mass retrievals of specific water-quality parameter codes. Therefore, a method for mass data retrievals of paired discharge (streamflow) and concentration measurements was needed for regional and national studies. The data-compilation method developed in this study is a four-step process that includes use of three computer programs and a database application to provide the data necessary to estimate the ambient water quality in a format that facilitates analysis (fig. 3). These tools are briefly described in this report and are fully documented in a series of user manuals on the CD-ROM accompanying this report.

The first step of the data-compilation process (fig. 3) is to download paired water-quality-constituent concentration and water-discharge data from NWIS Web with the National Water Information System Wizard (NWiz version 1.0) developed for this project (appendix 1). The NWiz program can be run to retrieve sampling-site information and data for an individual water-quality-monitoring station or it can be run in batch mode to retrieve information and data for multiple stations. The program works by translating the user's input to structured query language (SQL) commands that are transmitted over the internet to retrieve data. The program has a utility for finding sampling-site information within a geographic area and provides the ability to view graphs and summary statistics for station data retrieved from NWIS Web. Inputs to the program include the current NWIS Web internet address, the target directory for downloaded files, USGS station numbers, and the NWIS parameter codes for the water-quality constituents of interest. Output from the program includes NWIS station information (header) files in a relational database (RDB) format, station data files in RDB format and in the Kendall-Theil Robust Line (KTRLine) input-file format (Granato, 2006). NWiz also produces a "MissedFiles.txt" file to indicate stations that were not properly downloaded, and a "MissesData.txt" file to indicate station files that do not contain the requested data. Downloading these data, if done on a national scale, commonly resulted in creation of more than 10,000 NWIS Web RDB format files that contain more than 60 megabytes of data. Detailed instructions, program installation files, and program source code for the NWiz program are available on the CD-ROM accompanying this report (appendix 1).

Many of the files obtained from NWiz include either water-discharge or water-quality measurements, but not paired values. The second step of the data-compilation process (fig. 3) is to filter the results of a NWiz data retrieval to retain only those files containing paired data values. It was estimated that opening, visually inspecting, and deleting 10,000 NWIS RDB files manually would take 100 hours at a sustained average rate of almost 2 files per minute. Therefore, the National Water Information System Site Cleaner (NWISSC) (version 1.0) was written to automate this process (appendix 2). The NWISSC program automatically processes files in a directory (file folder) containing the results of NWiz water-quality data queries from NWIS Web. The NWISSC program uses the NWiz site-identification file to identify all station numbers in the query and the NWiz missed-data file (MissedData.txt) to identify unsuccessful query results. NWISSC cleans up the target directory by reading station numbers in the NWiz missed-data file, looking for associated data files in the target directory, and deleting these data files. Finally, the NWISSC program populates the "Status" text field on the form to indicate the total number of station queries, the number of station files that have both water-discharge and water-quality measurements, the number of station files in the NWiz missed-data file, and the number of files that were downloaded by NWiz but deleted by NWISSC because they did not contain both water-discharge and water-quality data. The NWISSC program generates a file of station numbers that have both water-discharge and water-quality data so that this list of "good files" can be used for subsequent processing and analysis. In comparison to the manual processing effort, it is estimated that NWISSC would require about 10 minutes of user input and (depending upon the user's computer) about 2 hours of computer processing time to process 10,000 NWIS RDB files. Detailed instructions, program installation files, and program source code for the NWISSC program are available on the CD-ROM accompanying this report (appendix 2).

The third step of the data-compilation process (fig. 3) is to extract paired values of water-discharge and water-quality data from the NWIS Web RDB files and to format these data for storage in and retrieval from the data-mining database. The NWIS database contains many records that include paired values of water-discharge and individual water-quality parameters, but not every water-quality measurement has an associated water-discharge measurement. Although measurement of water-discharge data with collection of a water-quality sample is a recommended practice (Hem, 1992), not all sampling programs require such paired data values. Instantaneous water-discharge measurements may be recorded without concurrent water-quality-sample collection; or if concurrent samples are collected, an individual water-quality constituent of interest may not have been included in the sampling effort at a given site. Data retrievals for a specific constituent may include many instances of independent water-discharge and water-quality measurements as well as paired values. If each file could be manually opened, searched, edited, copied, and the results of this process could be checked in an average of 1 minute per file, then processing 10,000 files would require about 167 hours of uninterrupted manual editing. The relational database file processor (RDBP version 1.0) was written to automate this process (appendix 3). By comparison, automatic processing of these data with RDBP is estimated to require about 10 minutes of user input and about 4 hours of computer-processing time to complete.

The RDBP program can be run for individual data files, or it can be run in batch mode for data files and sampling-site information (header) files. In data-processing mode, the program reads NWIS RDB and writes the paired values of water discharge and the selected water-quality parameter to an output file. Data-qualification codes are parsed from numeric values and stored in a separate column in the output file. The RDBP program has two output formats for data files. The default format is a tab-delimited format, in which each data element has its own column. This format is suitable for use with database, spreadsheet, statistical, and GIS applications. The secondary format is the KTRLine (Granato, 2006) input-file format, which has three tab-delimited columns (instantaneous water discharge, water-quality concentration, and sample metadata). The sampling-site information (header) file output is in a tab-delimited format suitable for use with many computer applications.

The RDBP program produces several diagnostic files. It produces an error file, if necessary, to alert the user about potential problems with input files. It produces a code file to document occurrence of water-discharge or water-quality values that have data-qualification codes in the NWIS Web RDB file. In batch mode, it produces a “good-station” file to document which data files have one or more paired measurements of instantaneous water discharge and water quality. Detailed instructions, program installation files, and program source code for the RDBP program are available on the CD-ROM accompanying this report (appendix 3).

The fourth step of the data-compilation process (fig. 3) is to import data into the USGS-FHWA Surface-Water-Quality Data Miner (SWQDM version 1.0) database (appendix 4). The SWQDM database is a relational database that can be used to associate the paired stream-water discharge and water-quality constituent-concentration data with an ecoregion or with any user-specified site location in the conterminous United States. All data that have been compiled and formatted in the first three compilation steps are associated with the water-quality-sampling stations that meet the selection criteria (fig. 3). The water-quality-data table (tblSurfaceWaterQuality) is designed to receive data in RDBP format. A station number code provides the linkage to sampling-station information, and a parameter code provides the linkage to information about the water-quality measurement. Detailed instructions and database installation files for the SWQDM database are available on the CD-ROM accompanying this report (appendix 4).

Data Interpretation

The interpretive methods used to generate regional planning-level estimates are based on two assumptions. The first assumption is that data from multiple stations in an ecoregion can be combined to develop a water-quality-transport curve that represents ambient water quality at unmonitored sites in that ecoregion. To develop regional water-quality-transport curves with data from multiple stations or to compare transport curves for different basins, it is necessary to use normalized water discharge (in cubic feet per second per square mile) so that the relation will reflect proportional water-discharge variations in data from each site (O’Connor, 1976; Glysson, 1987). The second assumption is that these transport curves, which are regression relations between the instantaneous discharge and concentration data available in NWIS, can be used for order-of-magnitude planning-level estimates of ambient EMC in receiving waters. These assumptions are necessitated by the scale of a national synthesis, the scope of this study, and limitations in available data. Furthermore, the data compilation and interpretation methods described in this report may be used with other information, such as local land-use data, for more selective regional or local data analysis.

The assumption that regional transport curves will provide planning-level estimates of ambient water quality in a region is not unprecedented. This assumption is based on the fact that the data and information, such as geology, physiography, vegetation, climate, soils, land use, and hydrology that are used to delineate ecoregions (Bailey, 1984; Omernik, 1995; Omernik and others, 2000; Shirazi and others, 2003) also are used as explanatory variables to indicate ambient water-quality characteristics (Langbein and Schumm, 1958; Syvitski and others, 2000; Shirazi and others, 2001; Griffith and others, 2002; Smith and others, 2003a; Simon and others, 2004). This assumption, however, does not imply that a single regional transport curve is capable of characterizing local variability for a specific site that may be caused by explanatory factors within the region. The assumption only implies that there may be differences among regions of the United States that may be characterized by development of separate transport curves. The ecoregion-level estimates developed herein are intended for an initial screening-level analysis, which may be followed by more detailed analysis with more site-specific data, if such an analysis is deemed necessary.

Instantaneous measurements of water discharge and constituent concentrations are used to develop planning-level estimates of ambient EMCs in receiving waters because relatively few data from streams and rivers are available to quantify EMCs. Concentrations of dissolved and suspended constituents commonly vary with discharge in hysteresis loops during a runoff event (O'Connor, 1976; Glysson, 1987; House and Warwick, 1998). Constituents characterized by washoff commonly exhibit higher instantaneous concentrations for a given discharge on the rising part of the hydrograph than on the falling part. Constituents from ground-water discharge, which are diluted by increasing runoff, commonly have lower concentrations for a given discharge on the rising part of the hydrograph than on the falling part. A transport curve fitted through instantaneous measurements of water discharge and constituent concentrations collected on the rising and falling parts of the hydrograph will, by definition, represent the central tendency of measured concentrations (Glysson, 1987; Helsel and Hirsch, 2002). Therefore, use of the event-mean flow as the explanatory variable in the transport curve for a given constituent will produce an estimate of the EMC for that constituent.

The first step in the data-interpretation process is to use the SWQDM database (appendix 4) to select data for analysis by ecoregion (fig. 3). For the current study, the site-selection and regionalization process yielded more than 24,000 surface-water-quality-monitoring stations in the conterminous United States. The distribution of sites, however, is not uniform among the ecoregions (fig. 4), and several ecoregions may have few, if any, data for a selected constituent. If sufficient data are not available in an ecoregion, or if a refined site-specific analysis is desired, the SWQDM database provides the option of locating data by geographic proximity to a user-defined study location by use of the Haversine Method (Sinnott, 1984; Drexel University Math Forum, 1999). The SWQDM database produces a three-column tab-delimited text file formatted for use with the KTRLine software. Detailed instructions and database-installation files for the SWQDM database are available on the CD-ROM accompanying this report (appendix 4).

The second step in the data-interpretation process is to use the KTRLine software (Granato, 2006) to do the nonparametric regression analysis (fig. 3). The Kendall-Theil robust line is a nonparametric regression procedure that is well suited for analysis of hydrologic data (Helsel and Hirsch, 2002). Hydrologic data sets commonly have statistical properties, such as outliers and skewed distributions, that are not ideal for application of parametric statistical techniques. Hydrologic data sets may not meet the theoretical assumptions of the ordinary-least squares (OLS) regression models that are commonly used (Hirsch and others, 1982; Koch and Smillie, 1986; Hirsch and others, 1991; Helsel and Hirsch, 2002). For example, hydrologic measurements may be reported as being less than one or more detection limits or greater than one or more maximum reporting limits. These poorly defined data points are commonly high-leverage/high-influence points. High-leverage points are defined as data in the tails of the distribution of predictor variables that deviate substantially from the OLS regression line that would be calculated if the point were absent from the data. The results of OLS analysis are sensitive to assumptions made about such data points. By comparison, robust methods minimize the effect of assumptions about high-leverage/high-influence points on the determination of relations between variables because they are not sensitive to values at the tails of the sample distributions (Helsel and Hirsch, 2002). For example, numerical studies indicate that one outlier may cause the OLS method to break down, whereas the breakdown point for the Kendall Theil method is at 29 percent of the data (Dietz, 1987; Wilcox, 1998, 2001). Furthermore, OLS regression is not well suited to regional studies that include the combination and analysis of multiple data sets, because the properties of the mixed distribution may not be suitable for use with parametric techniques (Hirsch and others, 1991).

The KTRLine software (Granato, 2006) also has several advantages over commonly used regression software because nonparametric multisegment regression tools are not commonly available as an option in regression software. The KTRLine software provides methods to facilitate data transformation and specification of multisegment regression models. Methods to facilitate data transformation are important because relations between water discharge and water-quality constituent concentrations commonly follow a power-law rather than a linear trend (O'Connor, 1976; Glysson, 1987; Simon, 1989; Hirsch and others, 1991; Vogel and others, 2003; 2005). The ability to specify multisegment regression models is important to characterize different processes that occur with different flow regimes (fig. 2) (O'Connor, 1976; Glysson, 1987; Simon, 1989; Hirsch and others, 1991). In theory and in practice, extension of transport curves to the limits of and slightly beyond available data can easily result in unrealistic estimates of concentrations and loads. Practitioners, however, commonly use slight extensions (much less than a log-cycle) of transport curves until additional data become available (J.R. Gray, U.S. Geological Survey, written commun., 2006). Because the multisegment models generated by the KTRLine program are based on nonparametric statistics, the potential for gross overestimation or underestimation of water-quality constituent concentrations is minimized within the limits of available data.

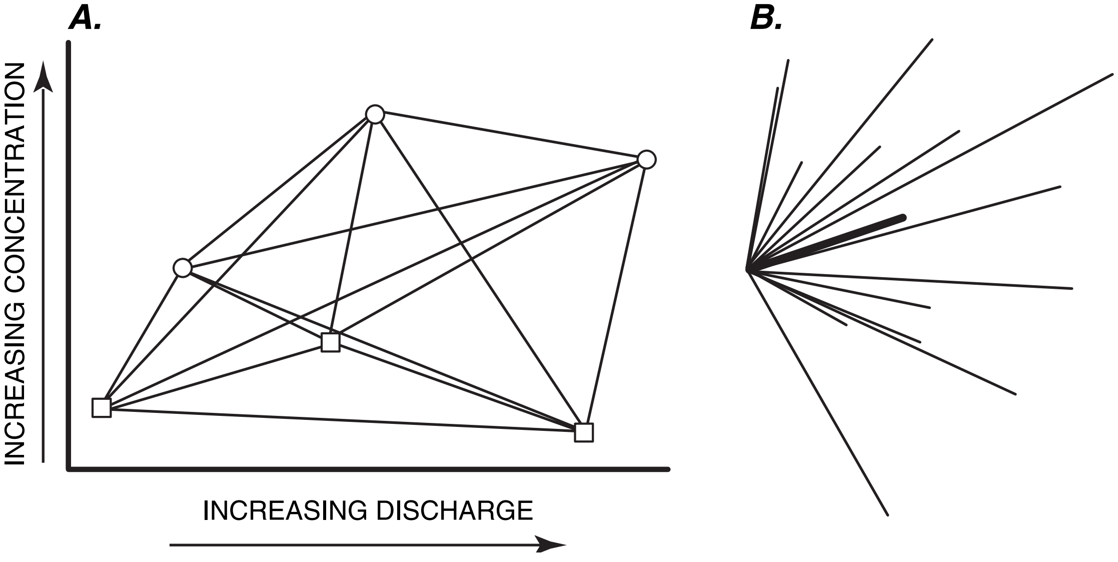


Figure 5. The manual method of determining the median slope: (*A*) all possible pairwise slopes between six data points, and (*B*) all possible slopes rearranged to meet at a common origin. The thick line is the median of the slopes. (Modified from Helsel and Hirsch, 2002.)

In the KTRLine software, the slope of the line is calculated as the median of all possible pairwise slopes between points (Theil, 1950; Sen, 1968; Helsel and Hirsch, 2002; Granato, 2006). Helsel and Hirsch (2002) provide a graphical example of the process (fig. 5). The intercept is calculated so that the line will run through the median of input data (Conover, 1980; Helsel and Hirsch, 2002; Granato, 2006). The KTRLine software also calculates the residual error, and the difference between the predicted value and the actual value for each data pair. For hydrologic data, these errors are related to measurement error (including sample collection, processing, and analysis) and natural variability caused by processes not evaluated in the regression model (Granato, 2006). Regression diagnostics such as the median error, the median absolute deviation (MAD), prediction error sum of squares (PRESS), the root mean square error (RMSE), the confidence interval for the slope, and the bias correction factor (BCF) for median estimates are calculated by use of nonparametric methods. Detailed instructions, program-installation files, and program-source code for the KTRLine software are available on the CD-ROM accompanying this report.

Standard methods for developing transport curves were used to develop regression models between normalized water discharge (discharge per unit of drainage area) and constituent concentrations (Glysson, 1987; Helsel and Hirsch, 2002). Use of these methods is integral to the design of the KTRLine software (Granato, 2006). The first step in the analysis is to examine the raw data graphically. In this study, a scatter plot of the data, a plot of residuals, and probability plots of the data and residuals were examined to assess the relations between variables and determine if a transformation was necessary to linearize the relation(s). Water discharge and concentration data sets showed substantial skew and non-constant variance (heteroscedasticity). Therefore, logarithmic transformation was used to improve linearity of relations between variables and to normalize regression residuals (Glysson, 1987; Helsel and Hirsch, 2002; Vogel and others, 2005). Base-10 logarithms were selected for use, rather than the natural logarithm, because the base-10 logarithms commonly are used in hydrologic applications (Stedinger and others, 1993).

Transformed data were examined and iteratively partitioned to find a one-, two-, or three-segment regression model. The objectives of the regression process were to define the simplest model that would

- visually and quantitatively provide a good fit to the data,
- be theoretically consistent with a dilution/washoff theory (for example see fig. 2),
- not produce apparently gross errors if extrapolated slightly beyond the range of available data,
- produce a median error that approaches zero,
- minimize heteroscedasticity in residuals with increasing water discharge, and
- normalize residuals.

In many cases, selection of a particular regression model to best meet these criteria was highly subjective, but confidence intervals for competing regression models commonly overlapped and residual error populations were comparable. The input data files, the output regression-results files, and Adobe PDF printouts of each step in the regression process are available in the regression-model directory on the CD-ROM accompanying this report.

The final step in the data-interpretation process (fig. 3) is the application of regression-model results. The KTRLine software (Granato, 2006) creates an output file that includes the equation and regression statistics for each linear segment of the regression model that is developed. Linear regression models are defined by the algebraic expression of a straight line. In a regression model, however, none of the terms are known, so each term must be estimated from available data, and an error term is added to account for individual departures from the estimated linear equation. Therefore, the equation for each linear segment of the regression model may be written as follows:

$$Y_i = m \cdot X_i + b + e_i \quad \text{for } i = 1 \text{ to } n \qquad (1)$$

where

X_i is the explanatory (independent, predictor, or X) variable for each data point (i);

Y_i is the response (dependent, predicted, or Y) variable for each data point (i);

m is the estimated slope;

b is the estimated intercept;

e_i is the residual error or uncertainty in the predicted Y-value for each data point (i); and

n is the number of paired XY data points in the sample that are characterized by a given segment.

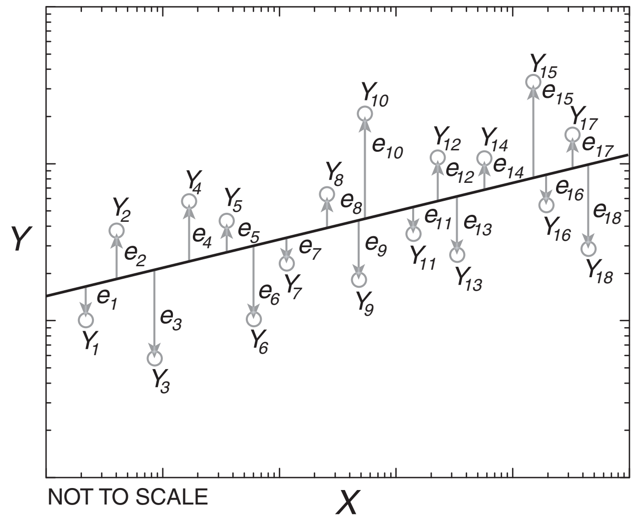


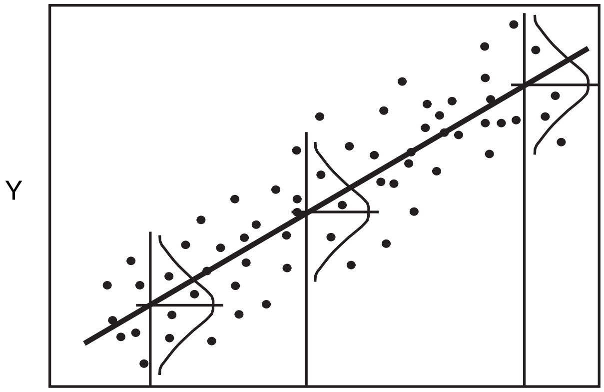
Figure 6. Schematic diagram showing the application of a regression model predicting Y from X. The variable e represents the random error component above or below the line for each of the 18 hypothetical XY data pairs.

The slope of the line (m) indicates the relation between water discharge and concentration. If the slope (m) of a segment of the regression line is approximately zero, this indicates that water quality does not vary with water discharge in the range of water discharge quantified by the segment. A positive slope indicates increasing concentrations with increasing water discharge, caused by runoff contribution of sediments and solutes (washoff). A negative slope indicates decreasing concentrations with increasing water discharge caused by dilution from precipitation and runoff.

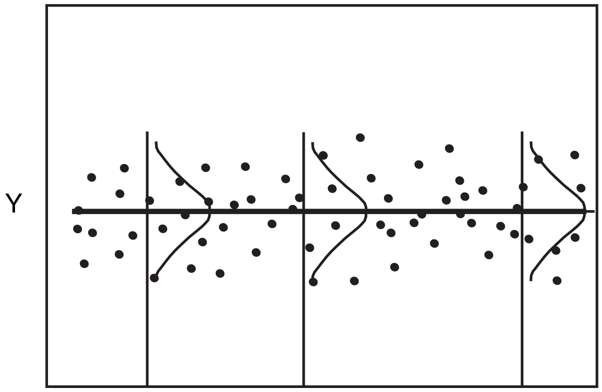
The estimated intercept (b) is the location parameter of the data in the interval represented by a given segment. For the Kendall-Theil line, the intercept is calculated with the paired medians of the X and Y data within the range used to define the segment. The intercept is the value of the line segment with slope m projected from the paired medians to the Y-axis (where X=0). If the slope is equal to zero, the value of the line will equal the median of Y data within the segment.

The error term (e_i) represents the variability in data measured vertically above and below the regression line (fig. 6). The variation in the error term is approximated by use of the MAD of the paired data set. The MAD is calculated as the median of the absolute values of the difference between each paired data point and the associated regression-prediction value (Helsel and Hirsch, 2002; Granato, 2006). The MAD statistic is about two-thirds of the standard deviation and about one-half of the interquartile range (IQR) for a population of e_i values with a normal distribution

A One-segment regression model with a positive slope



B One-segment regression model with a zero slope



C Two-segment regression model with a zero slope and a positive slope

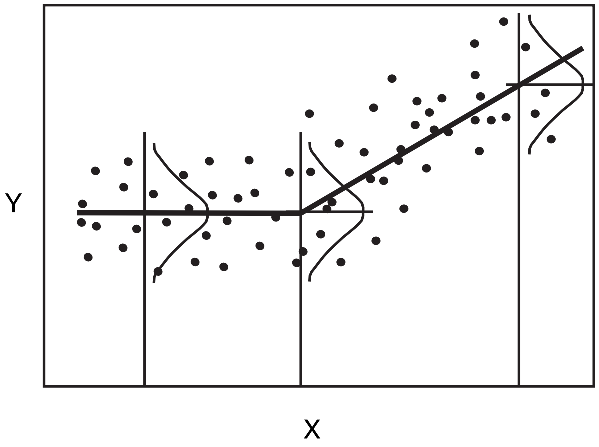


Figure 7. Normal distribution of data above and below the regression lines for *(A)* a one-segment regression model with positive slope, *(B)* a one-segment regression model with a zero slope, and *(C)* a two-segment regression model. (Modified from Riggs, 1968, p. 7.)

(Helsel and Hirsch, 2002). If the population of residuals is normally distributed, the regression-line prediction of the median Y_i value will approximate the average of Y_i values for a given X_i, and the MAD will approximate standard deviation of the Y_i values at that point within the interval represented by a given segment. If there is a linear one-segment relation between the predictor (X_i) and response (Y_i) variables, the slope of the line will be significantly different from zero (either positive, as in figure 7A, or negative) and the population of values will have a normal distribution of data above and below the line. Thus, these data vary with X and have a random error component. If the slope is not significantly different from zero (fig. 7B), the X_i term in equation 1 will drop out, the intercept of the line will represent the median and average of Y_i, and the population of values will have a normal distribution of data above and below the intercept. Thus, the data are random with respect to X and are described only by the statistics of the Y population, which are characterized by the intercept and error component of the regression analysis. If there are multiple processes that affect relations between the predictor (X_i) and response (Y_i) variables, there may be multiple relations that predominate over different ranges of the predictor variable (fig. 2). Figure 7C shows the case in which there is random variation in the response variable (for example base-flow conditions) until a second process predominates (for example runoff). In this case, each segment will have a different slope, intercept, and MAD.

The arithmetic average has special meaning in hydrologic studies designed to determine constituent loads because the arithmetic average constituent load is, in theory, the total mass of the constituent divided by the total amount of water that flows past the measurement point. Even a small bias in individual predictions can produce a substantial error in calculated loads. BCFs commonly are used, because the retransformed average of the transformed data does not equal the arithmetic average of the untransformed data (Koch and Smillie, 1986; Glysson, 1987; Gilroy and others, 1990; Crawford, 1991; Cohn and others, 1992; Hirsch and others, 1992; Helsel and Hirsch, 2002). The nonparametric smearing estimator (average of the retransformed log-regression residuals) proposed by Duan (1983) was selected for implementation with the KTRLine software because it performs reasonably well and is not sensitive to statistical assumptions of residual population characteristics (Koch and Smillie, 1986; Gilroy and others, 1990; Crawford, 1991; Hirsch and others, 1992; Helsel and Hirsch, 2002). The Duan (1983) BCF is called the smearing estimator because the method applies or “smears” the average retransformed error over all measurements.

The regression equations developed in this study were derived for stochastic data generation in runoff models. Bias correction for stochastic data generation is not necessary because application of the regression model with the random error component produces a population of individual estimates that may be retransformed and then averaged (Koch and Smillie, 1986). If the equations are used without the random error term to estimate the mass or loads of a constituent, then bias correction methods described by Granato (2006) are necessary for this purpose.

Regional Transport Curves for Selected Water-Quality Constituents

Regional transport curves that define relations between normalized water discharge (in cubic feet per second per square mile) and selected constituent concentrations can be used to provide planning-level estimates of surface-water quality in the conterminous United States. It is important to use concentration and water-discharge data that represent the full range of expected variability in both variables to derive water-quality-transport curves (Glysson, 1987). Several constituents were evaluated to assess the quantity and quality of available data in the different ecoregions of the conterminous United States. The slope of the transport curve indicates whether ambient constituent concentrations during stormflows are dominated by erosion and washoff, or by dilution. If the error component of a regression model is preserved, these models can be used for stochastic data generation in a Monte Carlo analysis of water-quality data.

Data Selection and Examination

Data for total nitrogen (parameter code p00600), total phosphorus (parameter code p00665), total hardness (parameter code p00900) and suspended sediment concentration (parameter code p80154) were selected, downloaded from NWIS Web, processed, and examined. The total number of paired water-discharge and water-quality samples is: total nitrogen—50,160; total phosphorus—246,403; total hardness—107,289; and suspended sediment—275,950. The objectives of the data-selection process were to obtain data for multiple sites in each ecoregion while minimizing the size of monitored drainage areas. Consideration of the sample-collection date may also be an important factor for interpreting regression results, if regression results reflect temporal rather than spatial differences in available data.

Nitrogen and phosphorus, which are essential nutrients for plant growth, can be harmful to aquatic ecosystems when natural or anthropogenic contributions cause excess concentrations of these constituents (Hem, 1992). Nitrogen is considered the limiting nutrient for many saltwater bodies, and phosphorus is considered the limiting nutrient for most freshwater bodies. Total nitrogen and total phosphorus were selected because the USEPA is currently developing water-quality criteria for these nutrients (Smith and others, 2003a; U.S. Environmental Protection Agency, 2005b), and because nutrients were addressed in the stream and lake analysis of the 1990 FHWA runoff-quality model (Driscoll and others, 1990a, b, c).

Total hardness was selected because it is used to calculate freshwater dissolved-metals criteria for trace metals such as cadmium, chromium, copper, lead, nickel, and zinc (U.S. Environmental Protection Agency, 2002). Total hardness also is used to estimate dissolved fractions of total-recoverable metal concentrations for cadmium and lead (U.S. Environmental Protection Agency, 2002). The 1990 FHWA runoff-quality model (Driscoll and others, 1990a, b) used a generalized national map of hardness concentrations that had intervals of about 60 milligrams per liter (mg/L) of hardness as calcium carbonate ($CaCO_3$). Estimates for total hardness concentrations are critical for receiving-water characterization because hardness concentrations determine both the CMC (acute criteria) and CCC (chronic criteria) values for trace metals in the receiving water. For example, relations between the U.S. Environmental Protection Agency (2002) freshwater dissolved CMC for zinc, lead, copper, and cadmium are shown as a function of total hardness concentration (as calcium carbonate) in figure 8. The U.S. Environmental Protection Agency (2002) tabulated values for a total hardness concentration of 100 mg/L, which is denoted as the solid, black vertical line in figure 8, and provided equations for calculating hardness-dependent CMCs in appendix B of the national recommended water-quality criteria. The figure also relates these criteria to common total hardness thresholds (soft, moderately hard, hard, or very hard) for describing water hardness (Hem, 1992). Although total hardness was selected for this ecoregion-scale analysis, trace elements were not selected because of substantial doubts about the veracity of much of the available historical data for measurements of fluvial trace-element concentrations (Smith and others, 1993; Breault and Granato, 2003).

Suspended sediment concentration was selected for several reasons. Primarily, it is considered one of the top water-quality problems in the United States (U.S. Environmental Protection Agency, 2005c). The USEPA is currently developing water-quality criteria for suspended and bedded sediments (U.S. Environmental Protection Agency, 2005c). Sediment also is the primary transport vector for trace elements (Breault and Granato, 2003) and organic contaminants (Lopes and Dionne, 2003). Local deposits of bed sediments deposited from runoff are associated with adverse ecological effects at some sites receiving highway runoff (Buckler and Granato, 2003). Many studies have used total suspended sediment (TSS) methods (parameter code p00530) to measure sediment, but TSS measurement methods are unreliable for quantification of fluvial-sediment concentrations (Gray and others, 2000) and for suspended sediment concentrations in runoff (Bent and others, 2003). TSS measurements commonly are biased low because the

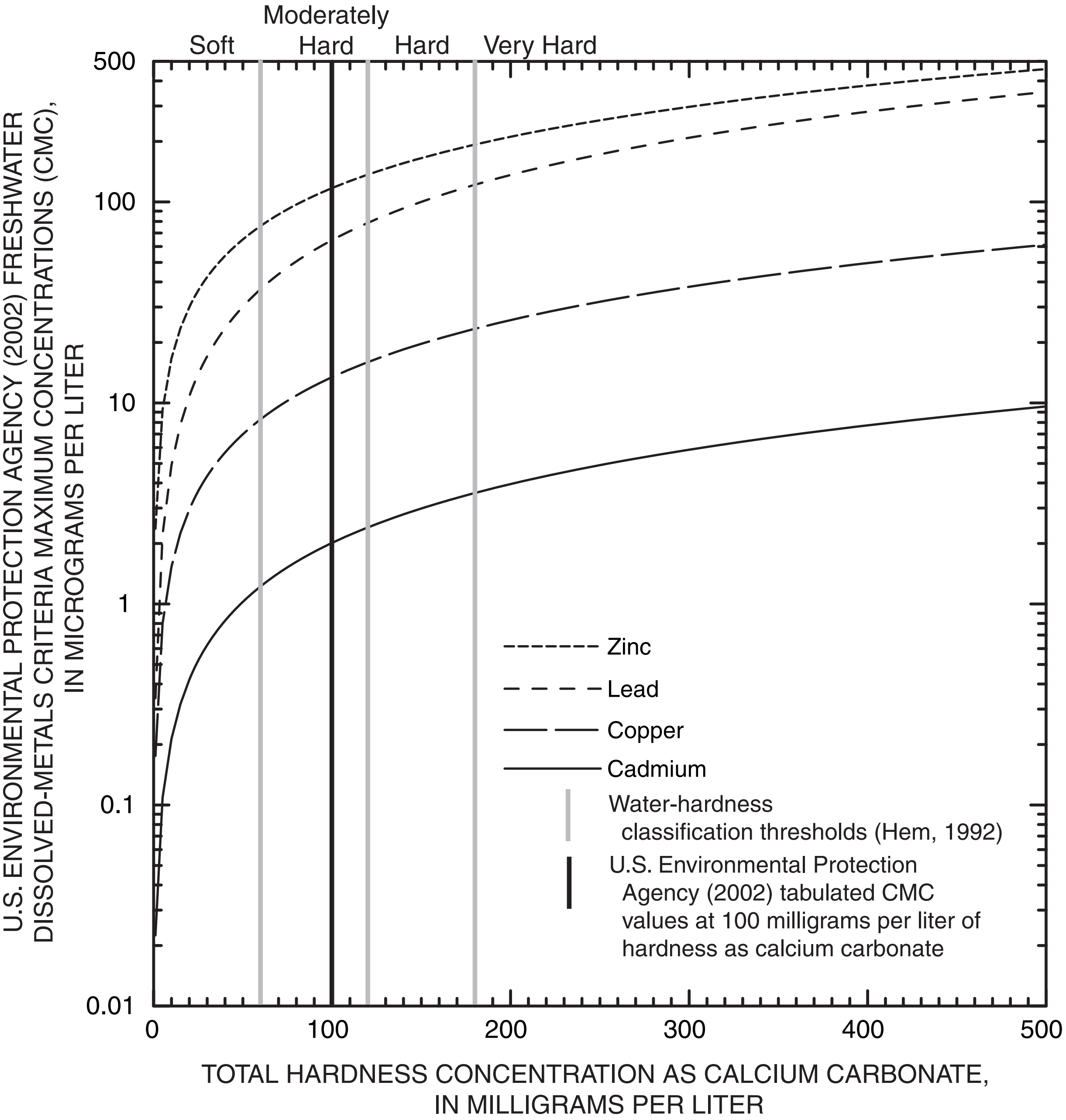


Figure 8. U.S. Environmental Protection Agency (2002) freshwater dissolved metals criteria maximum concentrations (CMC), in micrograms per liter as a function of total hardness concentration as calcium carbonate, in milligrams per liter. CMC values are shown in relation to the USEPA (2002) tabulated CMC values and in relation to commonly used total hardness thresholds (Hem, 1992).

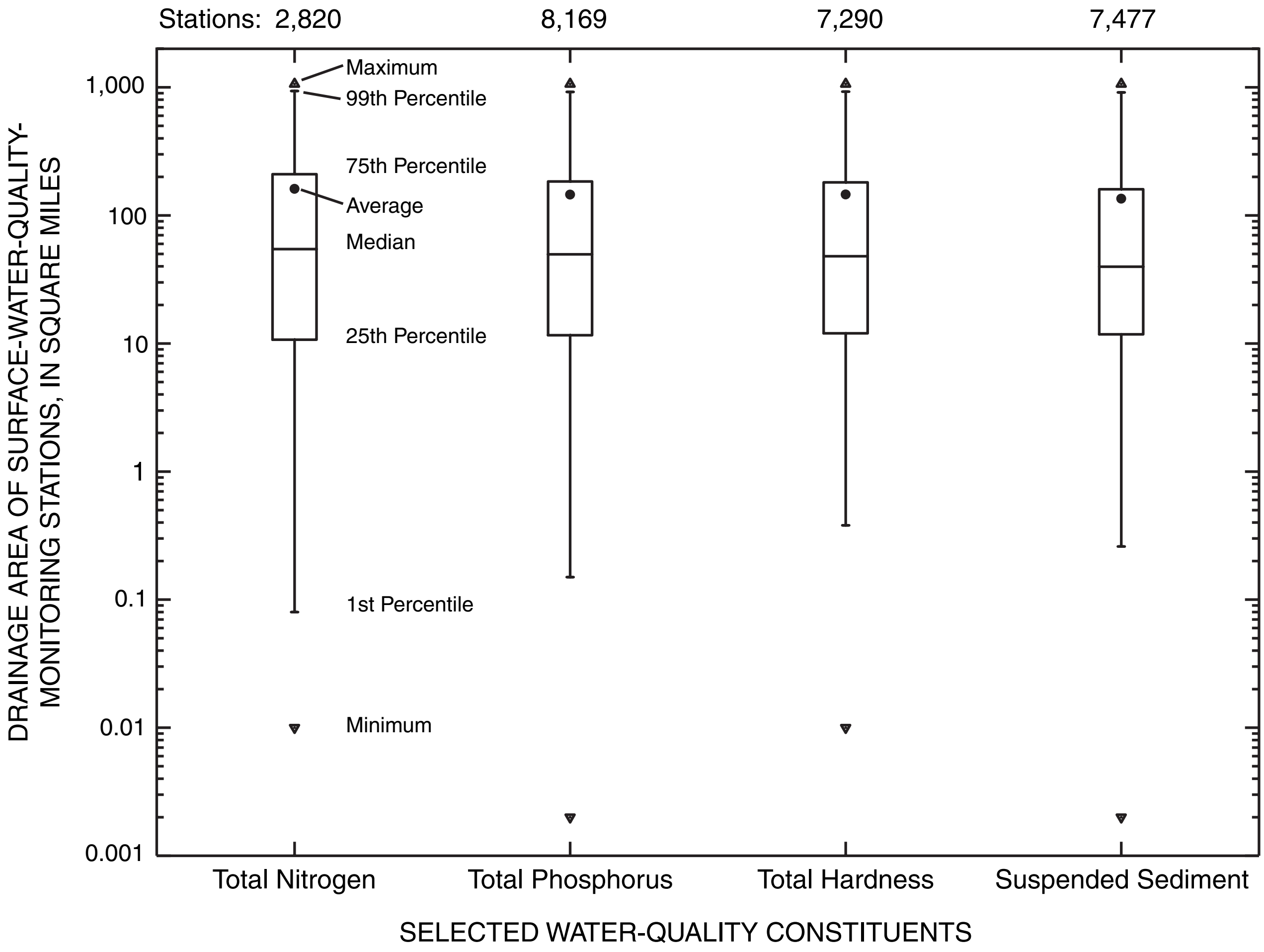


Figure 9. Distribution of surface-water-quality-monitoring station drainage areas for total nitrogen, water, unfiltered, milligrams per liter (p00600); phosphorus, water, unfiltered, milligrams per liter (p00665); hardness, water, unfiltered, milligrams per liter as calcium carbonate (p00900); and suspended sediment concentration, milligrams per liter (p8015) that were retrieved from the U.S. Geological Survey National Water Information System.

laboratory methods for measuring TSS tend to underrepresent all but the finer fractions of the suspended sediment (Gray and others, 2000).

The population distributions of drainage-basin sizes for the different water-quality constituents are relatively consistent even though there is a different number of stations with data for each constituent (fig. 9). The number of monitoring stations with paired water discharge and concentration data is: total nitrogen—2,820; total phosphorus—8,169; total hardness—7,290; and suspended sediment concentrations—7,477. About 25 percent of selected monitoring stations with data for each constituent have drainage basins that are less than about 10 mi^2. About 50 percent of the monitoring stations with data for each constituent have drainage areas from 10 to 200 mi^2. The maximum drainage area selected was 1,140 mi^2, and 99 percent of the monitoring stations have drainage areas less than 1,000 mi^2.

Populations of sampling dates vary considerably among the selected constituents (fig. 10). About 75 percent of total nitrogen samples were collected before 1981. Total nitrogen has the fewest number of stations and the fewest number of samples. Many surface-water-quality studies done since that time include analysis for different dissolved nitrogen constituents. By comparison, about 75 percent of the total phosphorus samples have been collected since 1980. About 50 percent of the total hardness data were collected from 1975 through 1982, probably because these data were needed to implement water-quality criteria for trace metals during initial implementation of the Clean Water Act (U.S. Environmental Protection Agency, 2003a). More than 50 percent of the suspended sediment data were collected after 1980. In this planning-level analysis, potential effects of temporal trends were assumed to be negligible for ecoregion-scale water-quality analysis. National water-quality trend analyses for a number of water-quality constituents indicated no detectable trends at almost 75 percent of monitored stations and a fairly even split between increasing and decreasing trends at the remaining stations for many constituents (Smith and others, 1987; Smith and others, 1993). More recent studies (for example, Smith and others, 1997) focus on local-scale explanatory variables, such as soils or land use, to examine spatial and (or) temporal variations in water quality. The data and tools (such as the SWQDM) provided in this study may be used to examine effects of such spatial or temporal factors at regional or local scales.

One exception considered in the current study is the effect of the Mount St. Helens eruption in 1980. Fallout from the eruption increased annual sediment yields in the northwestern United States by a factor of four or more and caused increased sediment concentrations in the immediate area for more than a decade (Meade and Parker, 1985; Smith and others, 1987; Dinehart, 1997; J.R. Gray, U.S. Geological Survey, written commun., 2005). Data characterizing the effects of this event were not used in the analysis because they do not reflect conditions of concern for most highway-runoff analyses.

The distribution of sampling dates, however, may be an issue for a more detailed analysis if water-quality trends substantially affect the representativeness of available data for current conditions. For example, in agricultural areas, it may be important that nitrogen fertilizer use in the U.S. peaked in 1981 and remained near the peak level throughout the 1990s (Smith and others, 1993; Mueller and Helsel, 1996). Similarly, phosphorus discharges to the environment peaked in the early 1980s and have been relatively constant since the early 1990s (Mueller and Helsel, 1996; Litke, 1999). Percentages of annual loads of total phosphorus from anthropogenic sources during the 1980s were: fertilizer application, 36 percent; manure application, 36 percent; and discharges from wastewater-treatment plants, 5 percent (Litke, 1999).

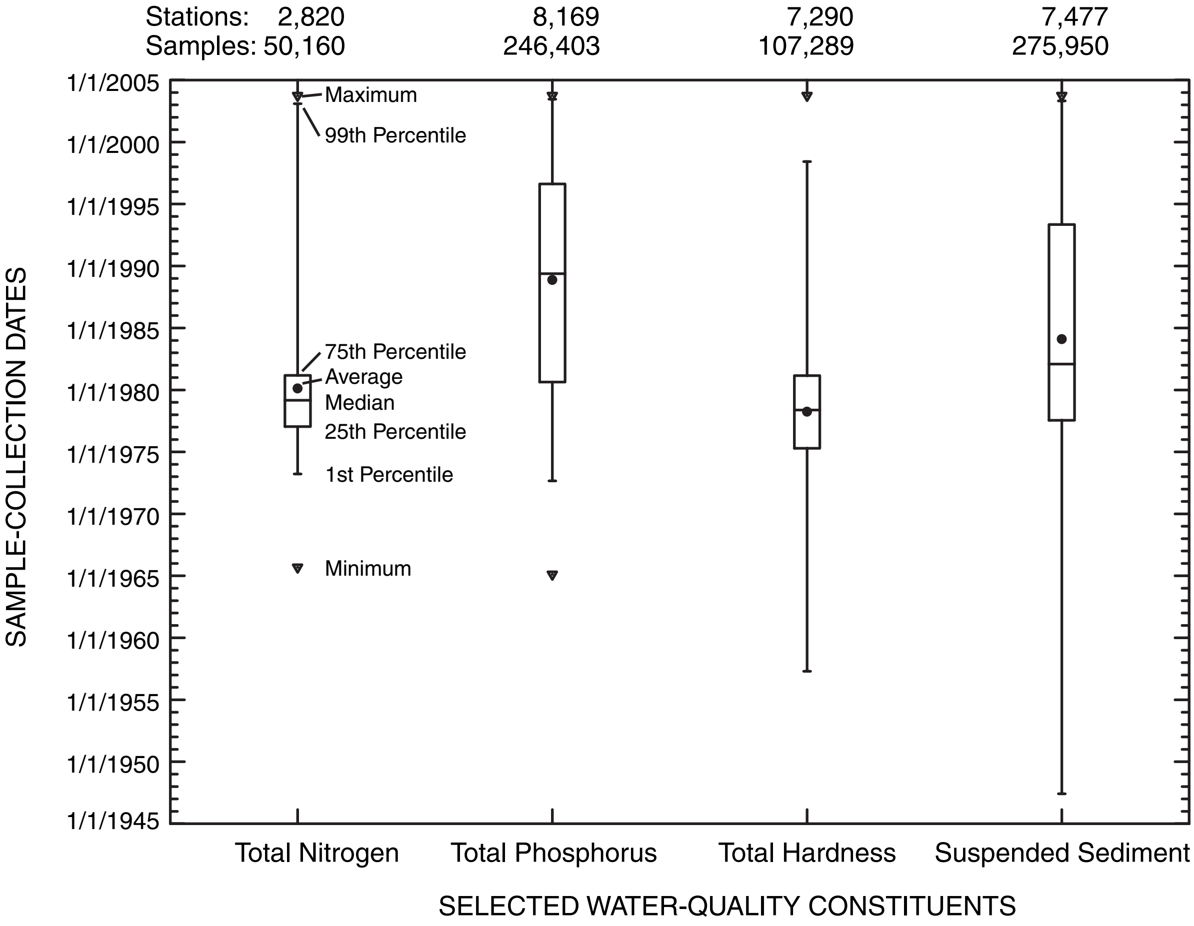


Figure 10. Distribution of water-quality sampling dates from 1945 through the end of water year 2003 for total nitrogen, water, unfiltered, milligrams per liter (p00600); phosphorus, water, unfiltered, milligrams per liter (p00665); hardness, water, unfiltered, milligrams per liter as calcium carbonate (p00900); and suspended sediment concentration, milligrams per liter (p8015) that were retrieved from the U.S. Geological Survey National Water Information System.

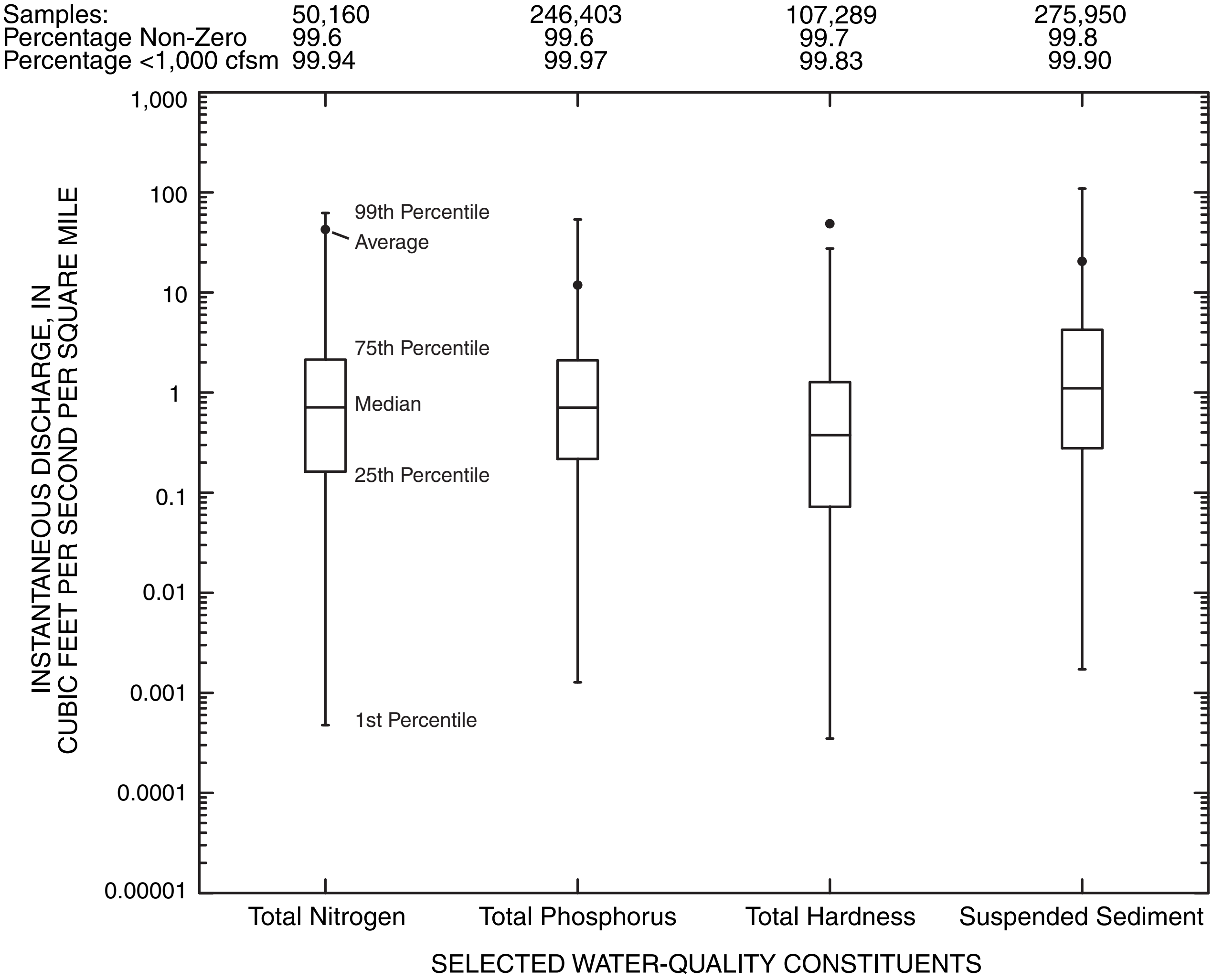


Figure 11. Distribution of non-zero instantaneous discharges in cubic feet per second per square mile (cfsm) measured with or estimated for surface-water-quality samples of total nitrogen, water, unfiltered, milligrams per liter (p00600); phosphorus, water, unfiltered, milligrams per liter (p00665); hardness, water, unfiltered, milligrams per liter as calcium carbonate (p00900); and suspended sediment concentration, milligrams per liter (p8015) that were retrieved from the U.S. Geological Survey National Water Information System. Non-zero stream discharges that are less than (<) 1,000 cfsm are shown.

Water-quality data from water-quality monitoring sites indicate decreasing concentrations of total phosphorus through time in many areas of the United States since the early 1980s. These decreases are probably caused by different control measures that were implemented to manage nutrients (Smith and others, 1993; Litke, 1999). It is expected that temporal trends in total hardness would be minimal because the calcium and magnesium comprising most of the measured hardness are primarily from geologic sources in many streams (Hem, 1992). Trends analysis indicates that improvements in soil-conservation practices throughout the United States have been effective in decreasing sediment concentration through time (Smith and others, 1993). Local changes in land use, however, may affect trends of all these constituents in selected receiving waters.

The ranges of instantaneous water-discharge measurements (p00061) associated with water-quality measurements are relatively consistent among the four selected water-quality constituents (fig. 11). About 98 percent of the instantaneous discharge values in the data set compiled for this study ranged from 0.001 to 50 cubic feet per second per square mile ($ft^3/s/mi^2$). About 0.3 percent of the data was associated with water discharges that were too small to measure, and about 0.9 percent of the data was associated with water discharges in excess of 1,000 $ft^3/s/mi^2$. The interquartile range of discharge measurements associated with the selected constituents was from about 0.2 to about 3 $ft^3/s/mi^2$. In theory, concentrations of water-quality constituents will approach asymptotic values (for example, see fig. 2) at high and low discharges that are characteristic of the physical and chemical properties of sources in a given basin. In reality, however, these asymptotic values are expressed as stochastic populations of values that reflect variations in precipitation, seasonality, land use, and other factors (O'Connor, 1976; Glysson, 1987; Simon, 1989; Hirsch and others, 1991). Therefore, it is necessary to have water-quality data from samples characterizing the full range of expected discharges at a given site. Nationally, the data set contains a substantial number of water-quality samples for more than five orders of magnitude of discharge normalized by drainage area (in cubic feet per second per square mile). Because the range of discharge is an important consideration in the application of transport curves, output from the KTRLine software includes the minimum and maximum values of discharge in the input data set, and the maximum value of discharge for each segment in a multisegment model.

Each of the selected water-quality constituents has a large range of measured concentrations (fig. 12). All four selected water-quality constituents have an interquartile range spanning a factor of five or more. About 98 percent of the concentration values in the data set compiled for this study ranged from 0.1 to 20 mg/L for total nitrogen, 0.005 to 5 mg/L for total phosphorus, 4 to 3,000 mg/L for total hardness, and 0.6 to 30,000 mg/L for suspended sediment concentration. The large ranges in total hardness and suspended sediment concentrations are indicative of large variations in regional water-quality values observed across the conterminous United States (Rainwater, 1962; Britton and others, 1983). Each of the four constituents has outliers that can be three to four orders of magnitude larger than the median value. Such outliers are not necessarily "bad data" but they may represent processes that are not of interest for evaluating the effects of highway runoff on receiving waters. For example, the maximum suspended sediment concentration value of 1,770,000 mg/L represents sediment transport from the Mount St. Helens explosion in 1980 (J.R. Gray, U.S. Geological Survey,

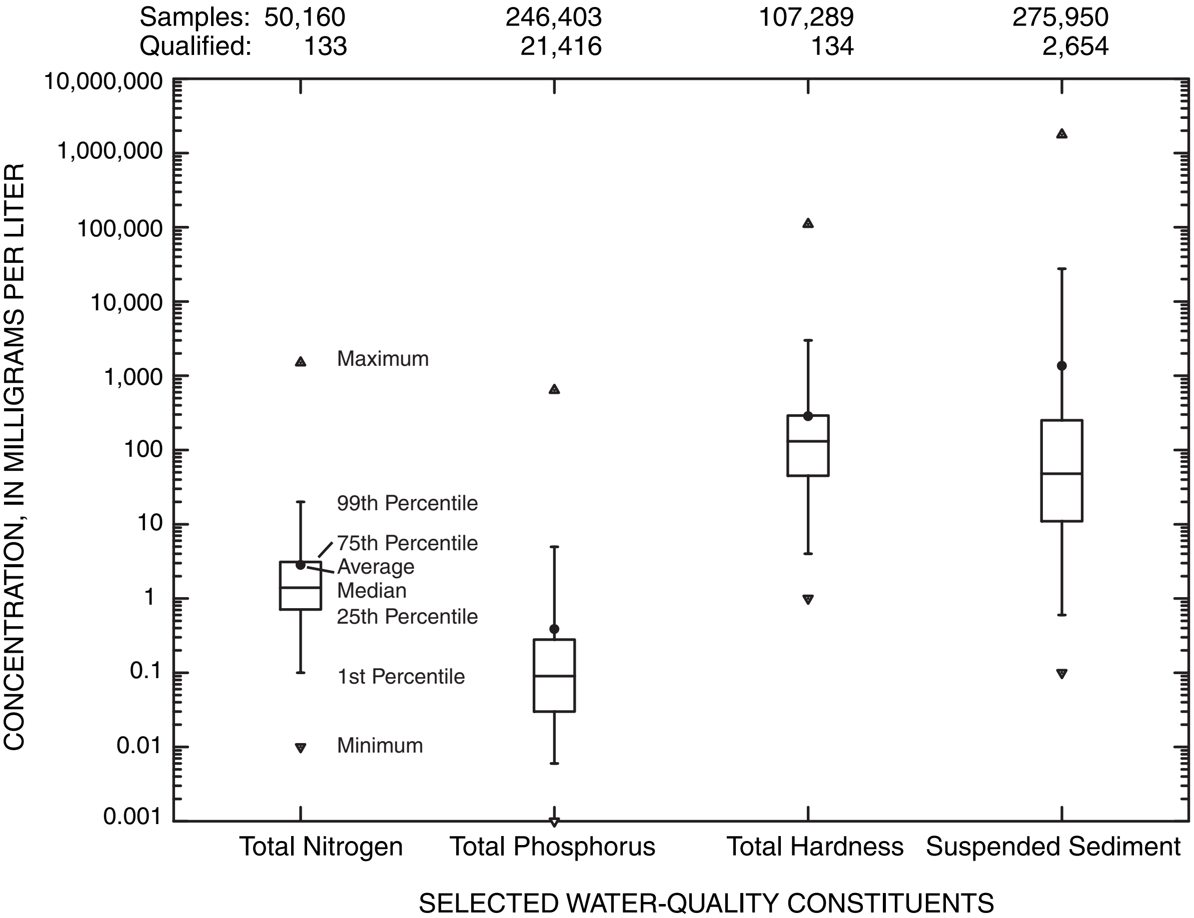


Figure 12. Distribution of measured concentrations of total nitrogen, water, unfiltered, milligrams per liter (p00600); phosphorus, water, unfiltered, milligrams per liter (p00665); hardness, water, unfiltered, milligrams per liter as calcium carbonate (p00900); and suspended sediment concentration, milligrams per liter (p8015) that were retrieved from the U.S. Geological Survey National Water Information System; qualified values include zeros, less thans, greater thans, estimated and altered values.

written commun., 2005). Some high values, however, may represent processes that affect runoff analysis. For example, values greater than 400,000 mg/L, which may be considered outliers in a national data set, represent hyperconcentrated flows that commonly occur in some arid regions of the southwestern United States (Beverage and Culbertson, 1964). These concentrations are of interest for evaluating the effects of highway runoff on receiving waters, although sediment concentrations from paved highways are expected to be much lower than concentrations commonly found in runoff from the surrounding arid lands (Langbein and Schumm, 1958).

Qualified values, which include estimated values, values less than detection limits, or values greater than maximum reporting limits, can be problematic for the interpretation of water-quality data (Helsel and Hirsch, 2002; Helsel, 2004). About 0.26 percent of total nitrogen values, about 9 percent of total phosphorus values, about 0.12 percent of total hardness values, and about 1 percent of suspended sediment concentration values are identified in NWIS as being qualified values (fig. 12).

Total nitrogen data were examined and included in the surface-water-quality data miner database (SWQDM), but were not included in the data used for regression analysis in this study. The total nitrogen data set included data from about 2,820 surface-water-quality monitoring stations, which is about 40 percent of the number of stations available for the other selected constituents (fig. 9). Total nitrogen data were not available for two ecoregions (number 76, the Southern Florida Coastal Plain; and number 79, the Madrean Archipelago in Arizona and New Mexico). Furthermore, 75 percent of the total nitrogen data were collected before 1981 (fig. 10), and total nitrogen data are expected to be affected by temporal trends (Smith and others, 1993; Mueller and Helsel, 1996). The available total nitrogen data set has about 18 to 50 percent of the measurements available for the other three data sets (fig. 12), and there are fewer than 100 paired values available for 29 percent of the 82 ecoregions that have total nitrogen data.

Regression Analysis

Regression equations for each of the 84 ecoregions in the conterminous United States were developed for total phosphorus (parameter code p00665), total hardness (parameter code p00900), and suspended sediment concentrations (parameter code p80154). A total of 252 regression models were specified for the three constituents selected for regression analysis. To document the analysis process, regression results including input files, output files, and regression-model development screens are recorded in the “Regress” directory on the CD-ROM accompanying this report. An example of a multisegment model development scenario is provided to familiarize the reader with methods used to formulate the regression models.

Results of the regression analysis are available in the “Regress” directory on the CD-ROM accompanying this report. The input data for each model are recorded in a tab-delimited text file named for the ecoregion and the parameter code. For example, the file “ER067p00900KT.txt” contains the total hardness (parameter code p00900) data for ecoregion 67, the Ridge and Valley ecoregion in the southeastern Appalachian Mountains (U.S. Environmental Protection Agency 2003b). Regression results are recorded in a text file named to identify the ecoregion (for example, 067RR.txt or SSC067RR.txt). Files for each water-quality constituent are grouped within each subdirectory (for example subdirectory “p00665TP” for total phosphorus). Adobe PDF files in these directories are printouts of selected regression-model development screens. The user is encouraged to view the original images in Adobe PDF format in the “Regress” directory of the accompanying CD-ROM. These file names ending in the letters “a” through “h” indicate the ecoregion and the content as follows:

a: The plot of the data and the regression line,

b: The probability plot of residuals,

c: The plot of transformed data and the regression line,

d: The plot of residuals versus the transformed discharge,

e: The probability plot of residuals from the transformed regression line, and (if a multisegment model was specified),

f: The plot of transformed data and the multisegment regression line,

g: The plot of residuals versus the transformed discharge, and

h: The probability plot of residuals from the transformed multisegment regression line.

For example, the name SSC067a.pdf is a plot of the suspended sediment concentration data and the regression line for ecoregion 67. The KTRLine software manual, which also is included on the CD-ROM accompanying this report, describes the input-file formats, regression methods, and output-file formats. An overview of the regression-model development process, KTRline screenshots, and the CD-ROM directory structure are illustrated on plate 1, which is also on the CD-ROM.

File names also indicate any modifications in the input data set. These modified input-data files are designated with an additional letter or number at the end of the root file name (for example, ER004p80154KTb.txt). If data from stations in the ecoregion were insufficient to support a multisegment regression analysis (about 20 or more data points), supplemental data were used from stations in neighboring ecoregions that were closest to the border between ecoregions. In other cases, there were too many (about 15,000) XY points to be processed within the limits of the KTRLine software (Granato, 2006). In these cases, data from the monitoring stations with the largest drainage basins were omitted to produce a manageable array of data from receiving waters most likely to be affected by runoff. In some cases, one far outlier (defined as data being beyond two orders of magnitude higher or lower than the nearest point and being of questionable physical veracity) was deleted from the input data set so the user could better view most of data on the KTRLine interface. In these cases, removing the outlier did not substantially affect the median-based statistics used to formulate the line. Finally, for two ecoregions, data were deleted from the input data set because they did not represent ambient conditions in the ecoregion. Sample results in ecogregion 2 (Puget Lowland) and ecoregion 4 (Cascades) characterizing sediment transport after the Mount St. Helens explosion were eliminated, which reduced maximum concentrations from 1,770,000 mg/L to 4,710 mg/L and from 1,060,000 mg/L to 11,600 mg/L, respectively.

Data for total hardness from ecoregion 67, the Ridge and Valley ecoregion in the southeastern Appalachian Mountains (U.S. Environmental Protection Agency 2003b), was selected to demonstrate the process of developing a multisegment regression model. This data set includes 4,057 paired water-discharge and total hardness measurements from samples collected at 265 stations from April 1959 to September 2003. The input data, the output file, and 15 regression-model development screen files documenting each step of the multisegment model-development process are recorded in the total hardness "p00900TH" subdirectory of the "Regress" directory on the CD-ROM accompanying this report. The KTRLine regression-model development screens for examination of a linear model, a model of the log-transformed data, and a multisegment model of log-transformed data, are shown on plate 1.

Once input and output files have been named by the user, the first step of the regression analysis is to examine a plot of the data and the initial untransformed regression line. Evaluation of the initial linear model (plate 1, screen 1) indicates that a transformation may be necessary because the scatter of untransformed data points looks like an exponential decay function rather than a linear relation between discharge and total hardness concentration. The initial regression-line estimate produces negative hardness concentrations for discharges greater than about 8 $ft^3/s/mi^2$. Probability plots of the X (water discharge) and Y (hardness concentration) data sets indicate skewed distributions (plate 1, screens 2 and 3). The plot of regression residuals in relation to discharge demonstrates nonconstant variance in the residuals of measured and predicted concentration with increasing discharge (plate 1, screen 4). The probability plot of residuals indicates a strongly skewed distribution (plate 1, screen 5). These factors indicate that a logarithmic transformation of both variables (commonly referred to as a log-log transform) may be necessary to formulate a linear regression model for these data (Glysson, 1987; Helsel and Hirsch, 2002; Vogel and others, 2005). The KTRLine software provides an interface for the user to individually transform the independent (discharge) and dependent (total hardness concentration) variables to reduce heteroscedasticity and to linearize data (plate 1, screen 6).

The next step is to examine a plot of the data and the initial regression line for the log-log transformed data. The KTRLine program graphs the data and calculates the regression equation in transformed space (plate 1, screen 7). Although there is substantial variability in this transformed data set, a general trend of decreasing hardness concentrations with increasing discharge is evident in most of the data. The probability plots of the logarithms of discharge and total hardness also indicate that transformation has reduced skewness of the input data sets (plate 1, screens 8 and 9). The plot of regression residuals in relation to the logarithm of discharge (plate 1, screen 10) does not show the nonconstant variance apparent in the untransformed regression model (plate 1, screen 4). Examination of the probability plot of transformed residuals reveals some skew in the residuals (plate 1, screen 11). The Kendall-Theil method, however, is not based on the assumption of normality in the input data or in the residuals (Helsel and Hirsch, 2002). Close examination of the plot of the transformed data in relation to the regression line, and the plot of the residuals in relation to the transformed discharge data shows that the one-line regression model overpredicts transformed hardness concentrations at low discharge values (in the range below the -0.5 value, about 0.32 $ft^3/s/mi^2$), and underpredicts transformed hardness concentrations at high discharge values (in the range above the 0.9 value, about 8 $ft^3/s/mi^2$). Furthermore, total hardness is a dissolved constituent that is expected to have a dilution pattern similar to the one shown in figure 2. These total hardness data demonstrate a relatively high base-flow concentration, dilution from increasing runoff, and a tendency to approach an asymptotic value at high discharge rates. Consequently, in this case, the three-line model provided the best fit to the data and was consistent with the theory of dilution of a dissolved ground-water source by precipitation-washoff described by O'Connor (1976).

The KTRLine software provides an interface for the user to specify multisegment regression models (plate 1, screen 12). The user may use even intervals of data, select breakpoint(s) by clicking data points on the graph, or entering the breakpoints in an input box. Specification of a multisegment model is an iterative process. The user chooses breakpoints to define two or more intervals of data, computes the regression lines by using the data in each interval, and then calculates the point at which two adjacent regression lines intersect (Glysson, 1987). The KTRLine software calculates each regression line with the data in the interval specified by the user and calculates the regression statistics for each line with the data in the interval specified by the intersection of adjacent regression lines (Granato, 2006). For some data sets, widely different models may be defined by selection of different breakpoints. Because the results of this analysis were intended for stochastic-data generation, one objective of the model-specification process was to find a model with a median deviation of zero and an approximately normal distribution of residuals. In this case, visual examination of the regression model (plate 1, screen 13), the plot of residuals with respect to discharge (plate 1, screen 14), and the probability plot of transformed residuals (plate 1, screen 15) all indicate a water-quality-transport curve that is representative of data over the complete range of sampled water-discharge values.

The final three-line regression model for total hardness concentrations in ecoregion 67 is shown on plate 1, screen 13. The model has a slope of zero, an intercept of about 2.04, and a MAD of about 0.15 for transformed discharges less than the -0.5 value. This indicates that total hardness varies randomly from about 40 to 300 mg/L, with a median value of about 110 mg/L as $CaCO_3$ for discharges less than about 0.32 $ft^3/s/mi^2$. The model also has a slope of zero for transformed discharges greater than the 0.9 value with an intercept of about 1.51, and a MAD of about 0.14. This indicates that total hardness varies randomly from about 12 to 85 mg/L, with a median value of about 32 mg/L as $CaCO_3$ for discharges greater than about 7.9 $ft^3/s/mi^2$. In comparison, the average total hardness concentration in precipitation in the region is about 0.28 mg/L as estimated from the 2003 U.S. National Atmospheric Deposition Program deposition isopleth maps for calcium and magnesium, (U.S. National Atmospheric Deposition Program, 2004). In theory, total hardness concentration would approach concentrations measured in precipitation as discharge increases. The minimum total hardness value measured in ecoregion 67 is 2 mg/L, indicating the importance of terrestrial sources relative to the precipitation input in this region. The interval between these segments has a slope of about -0.32 for transformed discharges, an intercept of about 1.9, and a MAD of about 0.18. This indicates that total hardness decreases exponentially as a function of water discharge in the range from 0.32 to about 7.9 $ft^3/s/mi^2$ with a median hardness concentration of about 80 mg/L with concentrations of total hardness at the median point ranging from about 23 to about 270 mg/L as $CaCO_3$.

If the plots of the residuals in relation to discharge and the probability plots of transformed residuals are compared for the one-segment model (plate 1, screens 10 and 11) and for the three-segment model (plate 1, screens 14 and 15), modest improvements in the fit of the data can be detected (plate 1). These differences are more easily seen on the page-size screen shots in the “Regress” directory on the CD-ROM accompanying this report. The differences between the one-line and three-line model predictions are especially important at low discharges. This is because a receiving stream with a relatively low discharge has a low dilution capacity for constituents in highway-runoff flows. The total hardness concentration at low discharge may be most critical for determining the potential for trace-metal toxicity in a receiving stream (U.S. Environmental Protection Agency, 2002).

This total hardness regression example provides a brief overview of the methods that were used to develop the water-quality-transport curves that are documented on the CD-ROM accompanying this report. More detailed information about regression and development of transport curves is presented by Glysson (1987), Helsel and Hirsch (2002), and other sources. Detailed instructions on the use of the KTRLine software are provided by Granato (2006). Specification of the regression models is a highly interpretive process with results that depend on the selection of the appropriate transformation, and for multisegment models, on the appropriate breakpoints between segments. Generally, the simplest model (the one with the fewest segments) that describes the data was selected; multisegment models were selected only if there was a substantial improvement in the statistical properties of the residual population. Three was the maximum number of segments used in these regression models to optimize the fit of the model and to maintain the theoretical assumptions describing the washoff or dilution processes involved.

Interpretation of Regression Results

Data from many ecoregions demonstrate relations between water discharge and water quality for the constituents analyzed (table 2). Only about 4 percent of the 252 regression models (the one-segment models with zero slopes) indicate that the relation between discharge and water quality is weak or nonexistent throughout the range of discharges associated with water-quality sample collection on an ecoregion scale (table 2). A one-segment model with zero slope indicates the population statistics for a water-quality constituent do not vary predictably with water discharge (fig. 7). Almost all the one-segment models with zero slope were for total phosphorus. Total phosphorus includes sediment-associated and dissolved species, which are commonly characterized by washoff and dilution, respectively. Because washoff would result in a positive relation between concentration and discharge, and dilution would result in a negative relation between concentration and discharge, a water-quality constituent that is strongly affected by both processes would not be expected to have a consistent relation with changes in discharge. The one-segment models with zero slope also may be an artifact caused by combining data from stations with differences in the dominant explanatory variables, such as different land uses, point sources, and natural characteristics. About 42 percent of the 226 multi-segment models have a zero slope within the interval of the first segment, indicating the random variation in constituent concentrations expected for base-flow conditions (figs. 2, 7). About 8 percent of the two-segment models have a zero slope within the interval of the second segment; this potentially indicates the random variation in constituent concentrations expected as concentrations approach an asymptotic value characteristic of fully mixed conditions (figs. 2, 7). Similarly, about 20 percent of the 48 three-segment models have a zero slope within the interval of the third segment.

Table 2. Summary of Kendall-Theil regression models of selected water-quality constituents and stream discharge indicating the type of models applied and the sign of slopes for different types of water-quality constituents.

[Regression coefficients and related statistics for each ecoregion are documented for each water-quality parameter in the Regress directory on the CD-ROM accompanying this report. --, not applicable]

Regression model		Sign of slope								
		First segment			Second segment			Third segment		
Type	Count	Negative	Zero	Positive	Negative	Zero	Positive	Negative	Zero	Positive
P00665 Phosphorus, water, unfiltered, milligrams per liter										
One Segment	10	1	9	0	--	--	--	--	--	--
Two Segments	57	7	35	15	1	7	49	--	--	--
Three Segments	17	4	9	4	1	0	16	3	3	11
P00900 Hardness, water, unfiltered, milligrams per liter as calcium carbonate										
One Segment	6	4	2	0	--	--	--	--	--	--
Two Segments	52	33	14	5	50	2	0	--	--	--
Three Segments	26	16	7	3	21	4	1	17	7	2
P80154 Suspended sediment concentration, milligrams per liter										
One Segment	10	0	0	10	--	--	--	--	--	--
Two Segments	69	8	26	35	0	0	69	--	--	--
Three Segments	5	0	3	2	0	0	5	0	0	5

Qualitative comparisons of the estimation capability of the regression models for total phosphorus, total hardness, and suspended sediment were made with various regional or national data summaries to assess how the regression-line estimates will represent order-of-magnitude planning-level estimates of receiving water quality throughout the conterminous United States. Comparisons are qualitative because the national-scale water-quality maps used are not available to the authors in a format that would facilitate more quantitative comparison. Also, national summaries commonly are highly interpretive and are based on extrapolation of available data. Furthermore, each study used in the comparison uses a different regionalization scheme. Most comparisons were made by using the median of water discharges associated with the water-quality samples for the region of interest to estimate a median concentration with the regression equation developed for that region. Estimates of suspended sediment concentrations using the transport curves developed in the current study were compared to regional estimates published by Simon and others (2004).

Total phosphorus has both a sediment-associated and a dissolved component. Both components have natural and anthropogenic sources, and are associated with both point and nonpoint sources (Smith and others, 1982; Hem, 1992; Smith and others, 2003a; Bricker, 2003). It is estimated, however, that about 95 percent of total phosphorus measured in natural waters are associated with suspended particulates (Hem, 1992). Results of the regression analyses for total phosphorus may reflect these complexities (table 2). Most of the one-segment models for total phosphorus have zero slopes, indicating random variation with increasing discharge (table 2). Many of the multiple-segment models have an initial slope of zero and a positive slope in the second or third segment. This pattern indicates random variations in concentration at base flow and increasing concentrations with stormflows, probably reflecting the predominance of the sediment-associated phosphorus species. All the total phosphorus regression models are recorded in the tab-delimited text file “P00665RR.txt” in the “Regress” directory on the CD-ROM accompanying this report.

Qualitative comparison of the regression results for total phosphorus by ecoregion with estimates of natural background concentrations of total phosphorus in headwater streams (Smith and others, 2003a) indicates that results from the Kendall-Theil regression equations provide good planning-level estimates of ambient stream-water quality. The ambient concentrations developed in the current study have a spatial variation that is similar to natural background concentrations predicted by Smith and others (2003a, b) who used data from a small subset of USGS stations that are minimally affected by anthropogenic inputs. For example, in both studies, concentrations less than about 0.05 mg/L are apparent in mountainous areas such as the Appalachians, the Boston

Mountains, the Coast Range, the Ouachita Mountains, the Ozark Highlands, the Sierra Nevada Mountains, and the Rockies. Both studies also indicate higher concentrations (greater than about 0.05 mg/L) in level III ecoregions in the Midwest that are affected by agriculture such as the Great Plains (Smith and others, 2003a, b). Total phosphorus concentrations from regression results from the current study, however, are generally greater than those predicted by Smith and others (2003a, b). This result probably is because the total phosphorus equations developed in the current study include the complete period of record for all available water-quality monitoring stations with drainage areas less than 1,040 mi^2 rather than the small subset of relatively natural reference stations used by Smith and others (2003a, b). These differences may reflect the difference between natural background water quality and ambient water quality at the ecoregion scale.

Total hardness is associated primarily with dissolved calcium and magnesium ions from ground water that discharges to surface waters (Hem, 1992). Negative slopes, indicating dilution with increasing discharge, predominate for the one-segment and the multisegment regression models (table 2). None of the one-segment models have a positive slope, about 5 percent of the segments in the two-segment models have a positive slope, and about 8 percent of the segments in the three-segment models have a positive slope. Each of the total hardness regression lines is recorded in the tab-delimited text file “P00900RR.txt” in the “Regress” directory on the CD-ROM accompanying this report. Qualitative comparison of the regression results by ecoregion with a national hardness map (by hydrologic-unit code areas) developed from National Stream Quality Accounting Network (NASQAN) data (Britton and others, 1983) indicates that the regression equations are useful for generating regional estimates of ambient water quality. For example, in this study and that of Britton and others (1983), areas such as the southern Rockies (ecoregion 21) and the Nebraska Sand Hills (ecoregion 44) are areas with lower hardness (less than 100 mg/L) than surrounding areas (greater than 200 mg/L). Large regional patterns, such as hardness concentrations less than 60 mg/L in the southeastern, northeastern, and northwestern United States and hardness concentrations greater than 1,000 mg/L in the Great Plains regions also are similar to the spatial patterns mapped by Britton and others (1983).

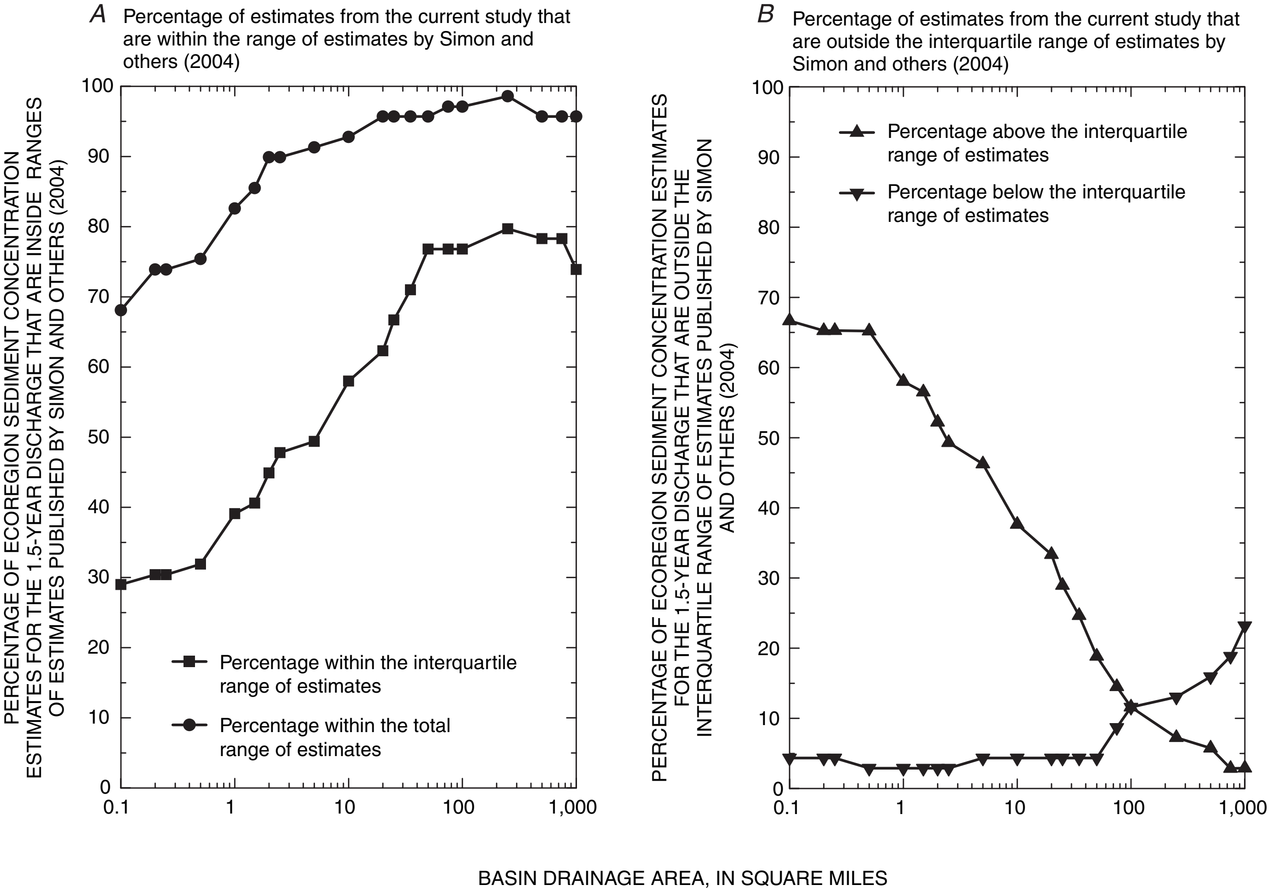
Suspended sediment is mobilized by washoff from the land surface and by erosion of streambeds and streambanks. Consequently, it is expected that as discharge increases, erosion and washoff will supply sediment to the stream and sediment concentrations will increase. For this study, all the regression models generated for suspended sediment concentration have one or more segments with positive slopes (table 2). These patterns indicate that sediment concentrations generally tend to increase with increasing discharge. Each of the suspended sediment concentration regression lines is recorded in the tab-delimited text file “P80154RR.txt” in the “Regress” directory on the CD-ROM accompanying this report. All one-segment models have a positive slope. All two-segment models have a positive slope in the second segment, and all three-segment models have positive slope in the second and third segments. About 12 percent of the 69 two-segment models have negative slopes in the first segment, but only 4 percent of these negative slopes are significantly different from zero within the calculated 95-percent confidence limit. Regression results for suspended sediment concentrations, therefore, follow theoretical relations with increasing discharge.

Comparison of the regression results for suspended sediment concentrations by ecoregion with three separate sources of information about patterns of suspended sediment concentration across the conterminous United States indicates that the results from the Kendall-Theil regression equations provide good planning-level estimates of ambient stream-water quality for this constituent. The ambient suspended sediment concentrations developed in the current study have a spatial variation similar to spatial variations shown on a national map of generalized suspended sediment concentrations in rivers (Rainwater, 1962). Spatial variations in both studies show the highest concentrations in the arid southwest and areas of the Great Plains and the lowest concentrations along the eastern seaboard. The concentrations published by Rainwater (1962) are greater than median concentrations estimated in the current study and are comparable to suspended sediment concentration estimates calculated by Simon and others (2004). Simon and others (2004) estimated water discharge and sediment concentrations for what they defined as a 1.5-year bankfull discharge condition, rather than for the average-annual estimates provided by Rainwater (1962). The ambient suspended sediment concentration estimates developed in the current study, however, have a similar spatial pattern and similar magnitude to annual average suspended sediment concentrations published by Britton and others (1983).

Estimates of suspended sediment concentrations developed from the KTRLine regression equations also are comparable to recent estimates of suspended sediment concentrations. Simon and others (2004) published regression equations predicting the 1.5-year bankfull discharge from drainage area for different ecoregions. They then used this estimated 1.5-year bankfull discharge value in conjunction with sediment-transport curves developed for each basin to estimate the associated suspended sediment concentration for each monitoring station in each ecoregion. Comparisons of suspended sediment concentration estimates developed from the KTRLine regression for each ecoregion were made by selecting a range of drainage areas, calculating the associated 1.5-year bankfull discharge by use of the discharge equation developed by Simon and others (2004). These estimates were compared to statistics for the 1.5-year sediment concentration estimates published by Simon and others (2004). These comparisons were completed only for ecoregions with data from five or more monitoring stations.

Results of the comparison indicate the utility of ecoregion-scale estimates for suspended sediment concentrations developed in the current study (fig. 13A). The percentage of ecoregion estimates for suspended sediment concentrations at the calculated 1.5-year bankfull discharge that fall within the range published by Simon and others (2004) is about 70 percent for a hypothetical drainage area of 0.1 mi^2. This percentage increases to an asymptotic value of about 95 percent of estimates if drainage areas greater than about 20 mi^2 are used to calculate the 1.5-year bankfull discharge and associated suspended sediment concentrations. Similarly, the percentage of sediment-concentration estimates within the interquartile range of the estimates published by Simon and others (2004) is about 30 percent if a drainage area of 0.1 mi^2 is used. This percentage increases to an asymptotic value of about 78 percent of estimates if drainage areas greater than 50 mi^2 are used (fig. 13A).

Examination of the percentage of sediment-concentration estimates that are either above or below the interquartile range of the estimates published by Simon and others (2004) indicates that the equations developed in the current study may overestimate sediment concentrations in basins with drainage areas less than about 50 mi^2 at the estimated bankfull-discharge conditions. The overestimates in sediment concentration, however, may be an artifact caused by differences in the drainage areas used by Simon and others (2004) and the drainage areas of stations used in the current study (fig. 13B). The patterns shown in figure 13 also may indicate uncertainties in the application of the equations developed by Simon and others (2004) that were used to estimate the 1.5-year bankfull discharge from drainage area. Any bias in estimated discharge would result in a systematic bias in suspended sediment concentrations estimated from a transport curve. The similar results of these two predictive approaches is notable, given the uncertainties in the representativeness of input data and the fact that 99 percent of daily discharges and concentrations are expected to be less than those for the 1.5-year bankfull-discharge event.


A Percentage of estimates from the current study that are within the range of estimates by Simon and others (2004)

B Percentage of estimates from the current study that are outside the interquartile range of estimates by Simon and others (2004)

Figure 13. Percentage of estimated sediment concentrations that are (*A*) within the range of estimates and (*B*) outside the interquartile range of estimates of sediment concentrations calculated from the estimated 1.5-year discharge as published by Simon and others (2004) for ecoregions in the conterminous United States.

Stochastic Data Generation

The water-quality-transport curves were developed for use with a stochastic data-generation algorithm to provide estimates of ambient concentrations from random estimates of discharge in receiving waters. Stochastic data-generation algorithms generate data values from a prescribed probability distribution and are used to estimate the probability of a given outcome. These algorithms can be used with water-quality-transport curves to provide planning-level estimates of ambient water quality from estimates of discharge in streams immediately upstream from potential highway-runoff sites. If this information is combined with stochastic estimates of runoff concentrations and flows from the highway, then the probability distribution of downstream constituent concentrations may be estimated. If such an analysis indicates the potential for adverse environmental effects in the receiving waters, the estimate may be refined with site-specific data and mitigation plans may be developed to reduce the potential for adverse effects.

The stochastic data-generation algorithm for generating concentration values is based on the regression equation expressed in equation 1. This algorithm is based on the assumption that the population of regression residuals (the error component, e_i) can be approximated as a random distribution centered on the regression line (fig. 7). If the regression equation is used with a stochastic data-generation algorithm, the error component term, e_i in equation 1, is estimated by use of a random-number generator designed to produce a standard normal variate. A standard normal variate is a member of population values that is normally distributed with a mean of zero and a standard deviation of one. The standard normal variate is then multiplied by the estimate of the standard deviation of the error population for the regression relation. If values of instantaneous concentration are predicted from instantaneous discharge, the standard deviation of the error population would equal about 1.5 times the MAD of regression residuals for the segment of the regression model associated with the discharge (X_i) used in equation 1 (Helsel and Hirsch, 2002). If EMCs are to be estimated from instantaneous measurements, the standard deviation of the error population may be assumed to be approximated by the MAD for planning-level estimates. The equation for e_i is

$$e_i = K_i \cdot \sigma, \tag{2}$$

where

e_i is the residual error or uncertainty in the concentration value predicted by the regression model for each data point (i);

K_i is a standard normal variate, obtained from a random-number generator for each data point (i); and

σ is the standard deviation of the error population, which is estimated from the MAD for the segment of the regression model associated with the discharge (X_i) used in equation 1. The combination of equations 1 and 2 yields an estimate for a random concentration (Y_i) predicted from a random discharge (X_i) as

$$Y_i = m \cdot X_i + b + K_i \cdot \sigma. \tag{3}$$

If the regression relation is formulated with the logarithms of concentration ($Y_i = \log10\ y_i$) and discharge ($X_i = \log10\ x_i$), then the relation can be expressed in real numbers as

$$y_i = x_i^m \cdot 10^{(b+Ki \cdot \sigma)}. \tag{4}$$

If a transformation is used to develop the regression model, a normal distribution of residuals will be generated within the transformed space, and the data will be proportionally skewed when estimates are retransformed to real space. If a random population of water discharges representing flow statistics during water-quality sample-collection efforts is used with equations 1 and 2, the result should be a random distribution of calculated concentrations that approximates the sampled population within the confidence intervals of the data collection and random data-generation processes.

Stochastic data generation with water-quality-transport curves may be used in a variety of ways. For example, if a single discharge value is used repeatedly with a set of (log) normal-random variates, this would produce a population of values centered on the value of the regression equation for that discharge with a standard deviation of σ. If a random population of discharges representing flow statistics in the region of interest is used with equations 1 and 2, the result should be a random distribution of calculated concentrations that are scattered in a normal distribution above and below each segment in the water-quality-transport curve. A spreadsheet program or statistical software may be used to generate a random distribution of water discharges that follow a lognormal distribution (Vogel and others, 2003; 2005) or a log-Pearson type III distribution (Chow and others, 1988; Kirby, 1972). If these discharge values are used in the transport equation with a second set of (log)normal-random variates, the user may derive a random distribution of constituent concentrations above and below the transport curve. Alternatively, the user may apply the transport curves with a historical water discharge record from NWIS Web (U.S. Geological Survey, 2004a) to estimate a random, historical population of constituent concentrations at a site in a given region. These estimates should be representative of ambient water quality in the region of interest.

The three-segment regression model developed for ecoregion 67 (listed in the file P00900RR.txt in the “Regress” subdirectory on the CD-ROM accompanying this report) was applied with a rudimentary random normal-number generator to demonstrate stochastic data generation of concentration data. In this example (fig 14A), paired measurements of total hardness and discharge (solid gray dots) are analyzed to determine a three-segment water-quality-transport curve (solid black line) with the KTRLine program (Granato, 2006). The total hardness data set for ecoregion 67 include 4,057 paired measurements of instantaneous discharge and total hardness concentrations. The boxplot along the *X*-axis of figure 14A indicates that samples were collected over more than four orders of magnitude of discharge. The different segments in the three-line model also indicate that the statistical properties of a water-quality data set may depend on the discharges at which samples were collected (fig. 14A). If samples are not collected to characterize variations over the full range of discharges, regression-line estimates and residual statistics may not quantify the range of concentrations that can be found in a region.

If an estimate of the water quality at an unmonitored site in the ecoregion is needed, this regional water-quality-transport curve can be used with regional flow statistics to estimate a population of concentration values for the site of interest (fig 14B). In this case, 1,000 discharge values were estimated in a lognormal distribution from regional stream discharge statistics, which are characterized by the boxplot along the *X*-axis of figure 14B. Application of the water-quality-transport curve (solid black line) and a lognormal distribution of residuals, generated using the MAD value for each applicable segment, yields a population of total hardness estimates (open circles) above and below the line (eq. 3). Comparison of the regression line, the stochastic data estimates, and the data from ecoregion 67 indicates that use of this total hardness transport curve with a stochastic component is sufficient for generating planning-level estimates of EMC values that are similar to most of the ambient water-quality measurements. These estimates may be sufficient for an initial environmental assessment done with a regional perspective, but estimates of regression statistics with data that are representative of a particular site may reduce uncertainty in the applicability of regression results. It should be noted that the regional transport curves developed in this study do not characterize site-to-site variability, seasonality, or trends through time in concentration or water discharge.

Total hardness is used in this example because variations in total hardness concentrations with discharge may affect the application of water-quality criteria (U.S. Environmental Protection Agency, 2002) for the hardness-dependent trace elements (fig. 8). For example, the calculated CMC (or acute criteria) and CCC (or chronic criteria) for copper are calculated from total hardness concentration values by use of log-linear relations (U.S. Environmental Protection Agency, 2002). The water-quality-transport curve for total hardness may be used to estimate associated CMC values for individual events (fig. 15). For example, if a normalized discharge of 0.37 $ft^3/s/mi^2$ is selected, the transport-curve value is 110 mg/L of total hardness (line 1a on fig. 15). If this value of total hardness is projected (line 1b) onto the log-linear relation between total hardness and the CMC for dissolved copper, the CMC value is about 14.7 µg/L (line 1c). Similarly, selection of a normalized discharge of 16.7 $ft^2/s/mi^2$, yields a total hardness value of about 32.5 mg/L and a dissolved copper CMC of about 4.67 µg/L (lines 2a, b, and c on fig. 15). This information indicates that the CMCs may decrease by a factor of 3 between these discharge values. The potential dilution, however, increases by a factor of 45 between these discharge values. Random variation in total hardness concentrations may be an important factor in ecoregion 67 because total hardness concentrations and associated CMC values for copper can vary substantially at any given discharge value. For example, the random values of hardness varies from almost 400 to about 30 mg/L of total hardness for discharges below 1.0 $ft^2/s/mi^2$; therefore, the corresponding CMC values for copper could range from about 50 to about 4 µg/L at any given discharge in this range.

Stochastic data analysis can provide information about the probabilities of different combinations of water discharge and water quality necessary to assess the potential need for and potential effectiveness of water-quality mitigation strategies. Use of a simple method, based on average values, may not reflect conditions that may be found at a given site because the prestorm base-flow volumes, upstream-runoff volumes, highway-runoff volumes, upstream water quality, and highway-runoff quality all may vary substantially from runoff event to runoff event. For example, the probability of exceeding a water-quality criteria for dissolved copper depends on the sources and distribution of copper in the receiving water and in highway runoff draining to a stream.

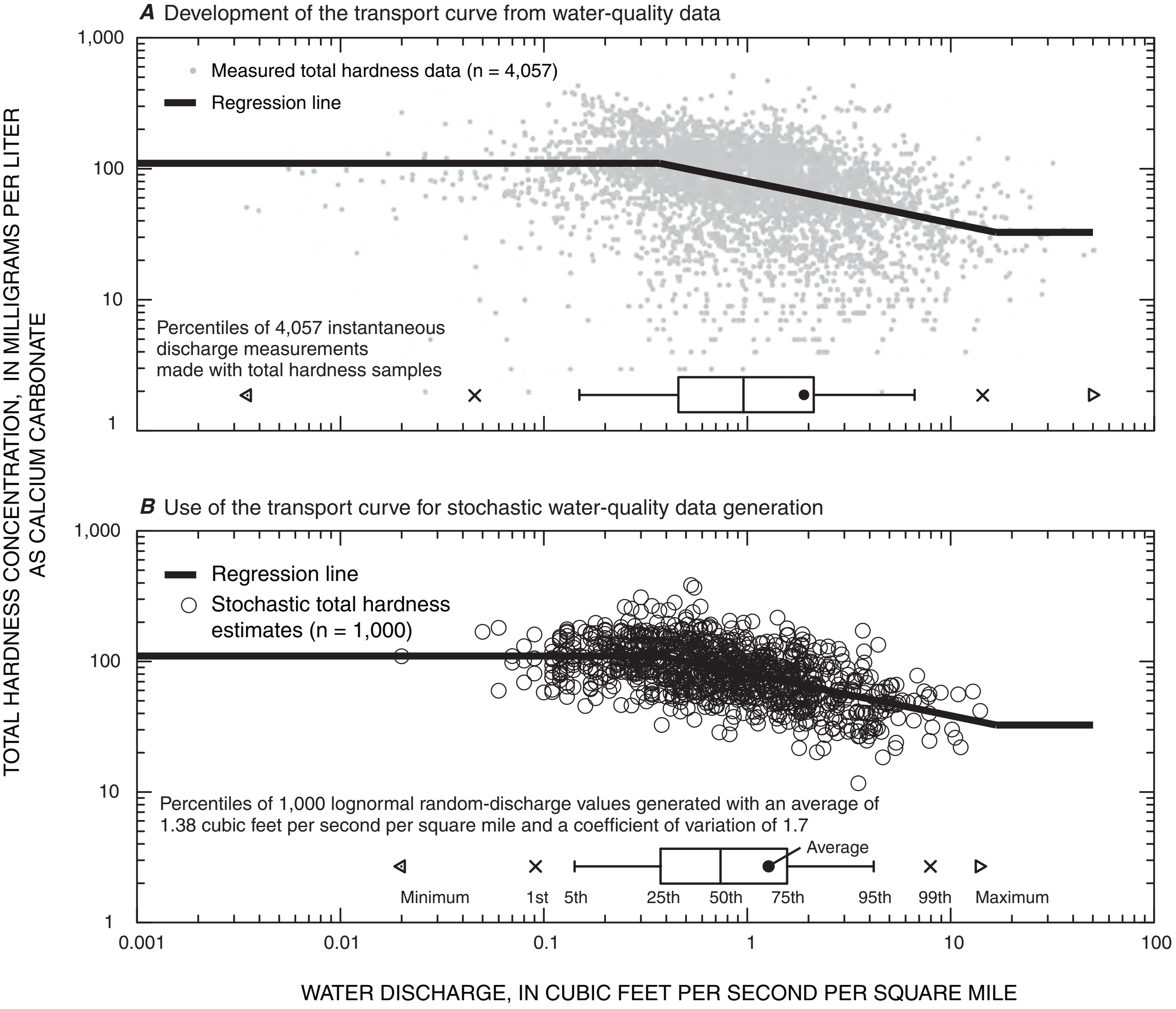


Figure 14. Example of *(A)* development and *(B)* use of a three-segment constituent-concentration-transport curve for stochastic generation of total hardness concentration data from regional average daily-flow statistics by use of a random-error component based on the median absolute deviation of water-quality data in the range covered by each segment within ecoregion 67.

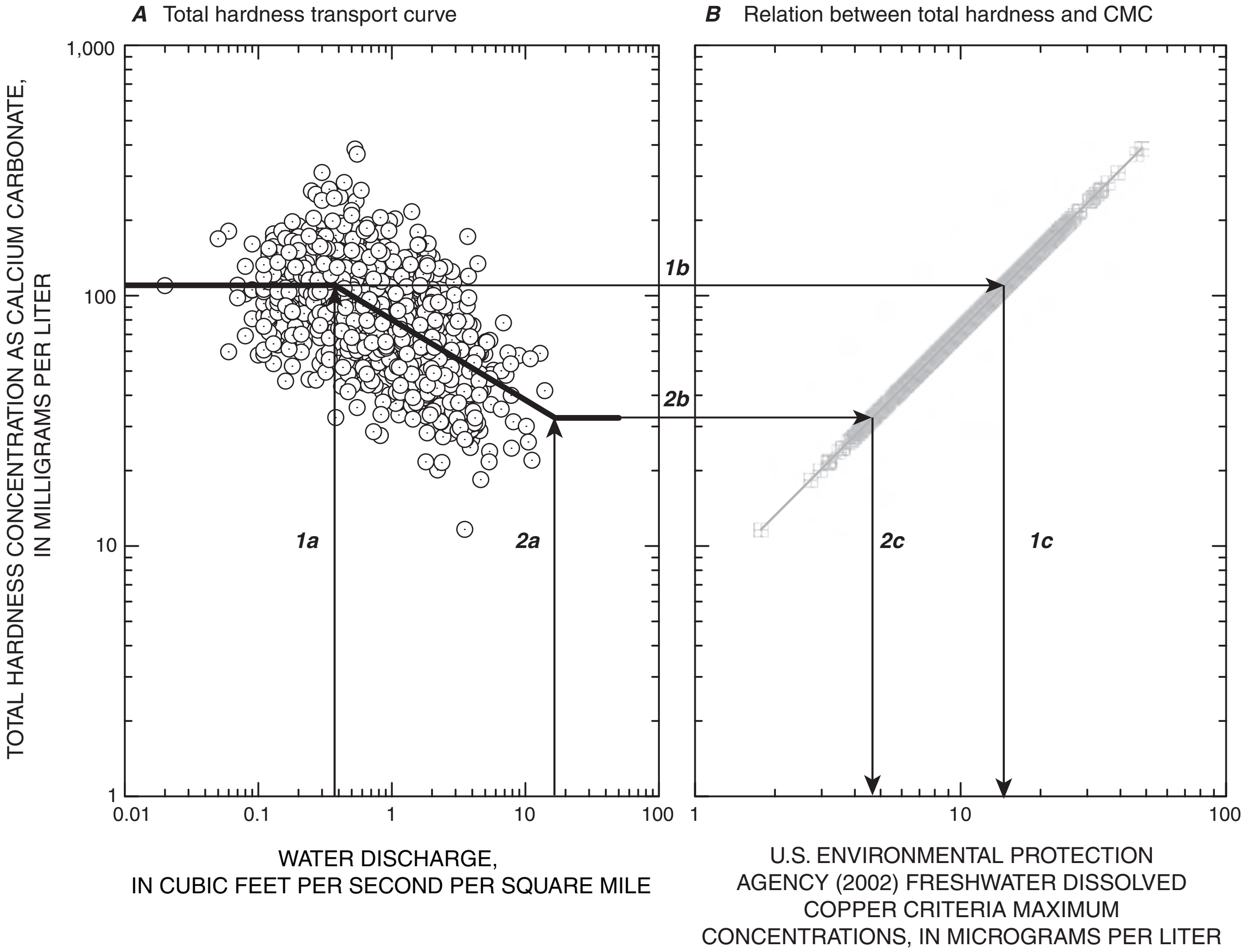


Figure 15. Freshwater dissolved copper criteria maximum concentrations (CMC) values calculated from *(A)* stochastic estimates of total hardness concentrations from the total hardness transport curve developed for ecoregion 67, and *(B)* the relation between total hardness and the U.S. Environmental Protection Agency (2002) CMC values.

Breault and Granato (2003) indicate that an increase in suspended sediment concentrations may increase the rate of and total amount of trace elements that are sorbed from solution onto the sediments. The transport-curve analysis for suspended sediment for ecoregion 67 (listed in the file P80154RR.txt in the "Regress" subdirectory on the CD-ROM accompanying this report) indicates that suspended sediment concentrations could increase by a factor of 20 (from about 7 to 140 mg/L) as discharge increased from 0.37 to 16.7 $ft^2/s/mi^2$ and suspended sediment concentrations show substantial variability at any given discharge. This variation in ambient sediment concentrations may have a profound effect on concentrations of dissolved copper in the receiving water body and would also vary stochastically from event to event.

Summary and Conclusions

Information and data about the magnitude and variation in ambient concentrations of chemical constituents representative of receiving-water quality can be used to evaluate highway runoff as a potential source of these constituents, to estimate the potential magnitude and frequency of adverse effects from runoff on receiving-water quality, and to evaluate the potential for best management practices to mitigate for potential adverse effects. Methods to develop robust estimates of water-quality-constituent concentrations in receiving waters as a function of water discharge (measured streamflow per unit time) can be used to develop and support runoff-quality models. These models quantify the risk of potential adverse effects from runoff as precipitation, discharge, ambient water quality, and highway-runoff quality vary from storm to storm. Concentrations of sediment, sediment-associated constituents (nutrients, organic compounds, and trace elements), and dissolved constituents in receiving waters commonly vary with discharge as a result of washoff and dilution processes. Regression relations between concentration and discharge, commonly known as water-quality-transport curves, are an accepted method for characterizing water-quality variation with discharge. Use of predictive transport curves is of particular interest because continuous records of daily water discharge are available at more than 20,000 streamflow-gaging stations across the United States for periods of years to decades, but records of measured constituent concentrations are comparatively sparse.

The U.S. Geological Survey and the U.S. Department of Transportation Federal Highway Administration did this cooperative study as part of a larger effort to develop a stochastic water-quality model. The stochastic water-quality model is designed to generate planning-level estimates of highway-runoff constituent concentrations and loads and the potential effects of these loads on receiving-water quality. This report documents the compilation and analysis of a national data set of paired values of instantaneous discharge and selected water-quality constituents to develop water-quality-transport curves for order-of-magnitude planning-level estimates of receiving-water quality at unmonitored sites upstream of highway-runoff outfalls in the conterminous United States.

Data from across the conterminous United States were grouped by U.S. Environmental Protection Agency, (USEPA) level III nutrient ecoregions to develop regional transport curves. These 84 ecoregions were selected because this system provides a consistent national context for regionalization of environmental data, and many of the characteristics used to define ecoregions also are commonly used as explanatory variables for water-quality analysis. A total of 24,581 surface-water-quality-monitoring stations throughout the conterminous United States were identified and cataloged for retrieval of water-quality data from the National Water Information System (NWIS) Web. The number of stations with paired water-discharge and concentration data is: total phosphorus—8,169; total hardness—7,290; and suspended sediment—7,477. Several water-quality constituents (total phosphorus, suspended sediment, and total hardness) were selected for analysis on the basis of the perceived importance for current water-quality analysis and potential availability of data sets from water-quality-monitoring stations in the 84 ecoregions. The total number of paired water-discharge and water-quality samples is: total phosphorus—246,403; total hardness—107,289; and suspended sediment—275,950. About 98 percent of the instantaneous stream-discharge values in the data set range from 0.001 to 50 cubic feet per second per square mile. About 98 percent of the concentration values in the data set compiled for this study range from 0.005 to 5 milligrams per liter (mg/L) for total phosphorus; 4 to 3,000 mg/L for total hardness; and 0.6 to 30,000 mg/L for suspended sediment concentrations.

Water-quality-transport curves for each of the 84 USEPA level III nutrient ecoregions in the conterminous United States were developed for total phosphorus, total hardness, and suspended sediment concentrations. The Kendall-Theil robust line (KTRLine) software (version 1.0) was used to develop water-quality-transport curves because this nonparametric method is resistant to the effects of outliers and nonnormality in residuals that characterize water-quality data. The KTRLine software can be used to develop multisegment regression models and provides the regression-residual statistics necessary for stochastic data generation. The objectives of the regression process were to define the simplest model that would visually and quantitatively provide a good fit to the data, be theoretically consistent with a dilution/washoff theory, not produce apparently gross errors if extrapolated slightly beyond the range of available data, produce a median error approaching zero, minimize heteroscedasticity in residuals with increasing water discharge, and normalize residuals. In each case, application of a random-number algorithm to the transport-curve equation with a stochastic error component produces a population of random concentrations that vary above and below each segment of the regression line.

A total of 252 regression models were specified for the three constituents selected for regression analysis. If the slope of a segment of the regression line is approximately zero, water quality does not vary with water discharge in the range of streamflow quantified by the segment. A positive slope indicates increasing concentrations with increasing water discharge (caused by a runoff or erosional contribution of sediments and solutes), and a negative slope indicates decreasing concentrations with increasing water discharge (dilution by precipitation and runoff). Only about 4 percent of the 252 regression models (the one-segment models with zero slopes) indicate that the relation between water discharge and water quality is weak or nonexistent throughout the range of sampled water discharges on an ecoregion scale. Almost all the one-segment models with zero slope were for total phosphorus. Total phosphorus includes sediment-associated and dissolved species, which are commonly characterized by washoff and dilution, respectively. The one-segment models with zero slope may be an artifact caused by combining data from stations with differences in the dominant explanatory variables such as different land uses, point sources, and natural characteristics. About 42 percent of the 226 multisegment models have a zero slope within the interval of the first segment; this indicates random variation in constituent concentrations expected for base-flow conditions. About 8 percent of the two-segment models have a zero slope within the interval of the second segment. This may indicate the random variation in constituent concentrations expected as concentrations approach an asymptotic value characteristic of fully mixed conditions. Similarly, about 20 percent of the 48 three-segment models have a zero slope within the interval of the third segment. The form and slopes of the models indicate that suspended sediment concentrations commonly increase with washoff and erosion, total hardness concentrations commonly are diluted by increasing streamflows, and total phosphorus concentrations are affected by both mechanisms.

The water-quality-transport curves developed in this analysis are intended for use with a stochastic data-generation algorithm. The algorithm is designed to provide order-of-magnitude planning-level estimates of water quality from estimates of discharge for stochastic analysis of ambient conditions immediately upstream of a potential highway-runoff site. The purpose of developing these stochastic data-generation algorithms for use with a highway-runoff model is to better quantify the potential for adverse environmental effects in the receiving waters. If such an analysis indicates the potential for adverse environmental effects in receiving waters, the estimate may be refined with site-specific data, and mitigation plans may be developed to reduce the potential for adverse effects. These data and analysis methods may be used to formulate regional curves for other constituents or to formulate more quantitative site-specific transport curves.

Semi-quantitative comparison of transport-curve estimates and published water-quality estimates for different areas of the United States demonstrates the utility of the transport-curve results. These comparisons indicate that the transport curves developed in this study are sufficient for generating planning-level estimates of water-quality-constituent concentrations in receiving waters upstream of highway-runoff sites. Use of these water-quality-transport curves with a random error component for stochastic data generation indicates that this method mimics the distribution of data in the ecoregion of interest. These models, however, are limited for quantitative site-specific application because of limits imposed by the spatial and temporal availability of data, the potential for temporal trends in water-quality data, local effects of physiography, land-use patterns, soil composition, and other factors.

Methods and tools for automated water-quality data compilation were developed to establish a preliminary database of selected water-quality constituents and to provide a means for future data-compilation efforts that may focus on a particular site or a particular water-quality issue. The data-compilation method developed in this study is a four-step process that includes use of three computer programs and a database application. The process allows the user to retrieve, filter, format, and select data for analysis. The interpretation process consists of selecting data, applying a nonparametric regression-analysis technique to the paired water-quality and discharge data, and applying the results of the regression analysis. The tools necessary to complete the process are described in this report and are fully documented in a series of user manuals on the CD-ROM accompanying this report. The data-compilation and interpretation methods described herein may be used with other information, such as local land-use data, for more selective regional or local data analysis. Results of the compilation and analysis are documented on the CD-ROM accompanying this report. The computer programs and database applications developed, utilized, and documented for the analysis also are fully documented on the CD-ROM accompanying this report. The CD-ROM also contains a data directory with more than 1,876,000 paired discharge and water-quality measurements that include 21 constituents commonly studied in highway- and urban-runoff studies.

Acknowledgments

The authors thank Gardner C. Bent, G. Douglas Glysson, and John R. Gray of the U.S. Geological Survey for the benefit of their experiences on the process for developing water-quality-transport curves. Dennis R. Helsel, Timothy A. Cohn, and Gary D. Tasker of the U.S. Geological Survey provided input and information on application of regression statistics. Patricia A. Cazenas of the Federal Highway Administration and Leslie DeSimone, John R. Gray, Jennifer Savoie, and Courtney Weed of the U.S. Geological Survey provided thoughtful and thorough technical reviews of this report, the software appendixes, and the associated CD-ROM.

References Cited

Albek, Erdem, 1999, Identification of the different sources of chlorides in streams by regression analysis using chloride-discharge relationships: Water Environment Research, v. 71, no. 7, p. 1310–1319.

Albek, Erdem, 2003, Estimation of point and diffuse contaminant loads to streams by non-parametric regression analysis of monitoring data: Water Air and Soil Pollution, v. 147, p. 229–2003.

Bailey, R.G., 1984, Testing an ecosystem regionalization: Journal of Environmental Management, v. 19, p. 239–248.

Barnwell, T.O. Jr., and Krenkel, P.A., 1982, Use of water quality models in management decision making: Water Science and Technology, v. 14, no. 9–11, p. 1095–1107.

Bent, G.C., Gray, J.R., Smith, K.P., and Glysson, G.D., 2003, A synopsis of technical issues for monitoring sediment in highway and urban runoff, *in* Granato, G.E., Zenone, C., and Cazenas, P.A. (eds.), National highway runoff water-quality data and methodology synthesis, Volume I—Technical issues for monitoring highway runoff and urban stormwater: Washington, D.C., U.S. Department of Transportation, Federal Highway Administration, FHWA-EP-03-054, p. 111–163.

Beverage, J.P., and Culbertson, J.K., 1964, Hyperconcentrations of suspended sediment: Journal of the Hydraulics Division, Proceedings of the American Society of Civil Engineers, v. 90, no. HY6, p. 117–129.

Breault, R.F., and Granato, G.E., 2003, A synopsis of technical issues for monitoring trace elements in highway and urban runoff, *in* Granato, G.E., Zenone, C., and Cazenas, P.A. (eds.), National highway runoff water-quality data and methodology synthesis, Volume I—Technical issues for monitoring highway runoff and urban stormwater: Washington, D.C., U.S. Department of Transportation, Federal Highway Administration, FHWA-EP-03-054, p. 165–233.

Bricker, O.P., 2003, An overview of the factors involved in evaluating the geochemical effects of highway runoff on the environment, *in* Granato, G.E., Zenone, C., and Cazenas, P.A. (eds.), National highway runoff water-quality data and methodology synthesis, Volume I—Technical issues for monitoring highway runoff and urban stormwater: Washington, D.C., U.S. Department of Transportation, Federal Highway Administration, FHWA-EP-03-054, p. 81–110.

Britton, L.J., Goddard, K.E., Briggs, J.C, 1983, Quality of rivers of the United States, 1976 water year—based on the National Stream Quality Accounting Network (NASQAN): U.S. Geological Survey Open-File Report 80–594, 432 p.

Brown, J.W. (ed.), 2006, Eco-logical—An ecosystem approach to developing infrastructure projects: Washington, D.C., U.S. Department of Transportation, Federal Highway Administration, FHWA-HEP-06-011, 96 p.

Buckler, D.R., and Granato, G.E., 2003, Assessing biological effects from highway-runoff constituents, *in* Granato, G.E., Zenone, C., and Cazenas, P.A. (eds.), National highway runoff water-quality data and methodology synthesis, Volume I—Technical issues for monitoring highway runoff and urban stormwater: Washington, D.C., U.S. Department of Transportation, Federal Highway Administration, FHWA-EP-03-054, p. 305–351.

Cazenas, P.A., Bank F.G., and Shrouds, J.M., 1996, Guidance on developing water-quality action plans: Washington, D.C., Department of Transportation, Federal Highway Administration, Office of Natural Environment, 9 p.

Chandler, R.D., 1994, Estimating annual urban nonpoint pollutant loads: Journal of Management in Engineering, v. 10, no. 6, p. 50–59.

Chow, V.T., 1971, Stochastic hydraulics—A challenging field of study: *in* Chiu, C.L. (ed.), Stochastic hydraulics: Pittsburgh, PA, University of Pittsburgh, School of Engineering Publication Series no. 4, p. 3–8.

Chow, V.T., Maidment, D.R., Mays, L.W., 1988, Applied hydrology: New York, McGraw-Hill Publishing Company, 572 p.

Clarke, R.T., 1990a, Statistical characteristics of some estimators of sediment and nutrient loadings: Water Resources Research, v. 26, no. 9, p. 2229–2233.

Clarke, R.T., 1990b, Bias and variance of some estimators of suspended sediment load: Hydrological Sciences Journal, v. 35, no. 3, p. 253–261.

Cohn, T.A., 1995, Recent advances in statistical methods for the estimation of sediment and nutrient transport in rivers: U.S. National Report to International Union of Geodesy and Geophysics 1991–1994, Reviews of Geophysics, Supplement, v. 33, p. 1117–1123.

Cohn, T.A., Caulder, D.L., Gilroy, E.J., Zynjuk, L.D., and Summers, R.M., 1992, The validity of a simple log-linear model for estimating fluvial constituent loads—An empirical study involving nutrient loads entering Chesapeake Bay: Water Resources Research, v. 25, no. 9, p. 2353–2363.

Conover, W. L., 1980, Practical nonparametric statistics, (2d ed.): New York, John Wiley and Sons, 493 p.

Crawford, C.G., 1991, Estimation of suspended-sediment rating curves and mean suspended-sediment loads: Journal of Hydrology, v. 129, p. 331–348.

Dietz, E.J., 1987, A comparison of robust estimators in simple linear regression: Communication in Statistics-Simulation, v. 16, p. 1209–1227.

Dinehart, R.L., 1997, Sediment transport at gaging stations near Mount St. Helens, Washington, 1980–90, data collection and analysis: U.S. Geological Survey Professional Paper 1573, 111 p.

DiToro, D.M., 1984, Probability model of stream quality due to runoff: Journal of Environmental Engineering, v. 110, no. 3, p. 607–628.

Drexel University Math Forum, 1999, Deriving the haversine formula: Accessed August 2, 2004 at *http://mathforum.org/library/drmath/view/51879.html*

Driscoll, E.D., Shelley, P.E., Gaboury, D.R., and Salhotra, Atul, 1989, A probabilistic methodology for analyzing water quality effects of urban runoff on rivers and streams: U.S. Environmental Protection Agency, Office of Water, EPA 841-R89-101, 128 p.

Driscoll, E.D., Shelley, P.E., and Strecker, E.W., 1990a, Pollutant loadings and impacts from highway stormwater runoff, Volume I—Design procedure: Federal Highway Administration Final Report FHWA-RD-88-006, 67 p.

Driscoll, E.D., Shelley, P.E., and Strecker, E.W., 1990b, Pollutant loadings and impacts from highway stormwater runoff, Volume II—User's guide for interactive computer implementation of design procedure: Federal Highway Administration Final Report FHWA-RD-88-007, 23 p.

Driscoll, E.D., Shelley, P.E., and Strecker, E.W., 1990c, Pollutant loadings and impacts from highway stormwater runoff, Volume III—Analytical investigation and research report: Federal Highway Administration Final Report FHWA-RD-88–008, 160 p.

Driscoll, E.D., Shelley, P.E., and Strecker, E.W., 1990d, Pollutant loadings and impacts from highway stormwater runoff, Volume IV—Research report data appendix: Federal Highway Administration Final Report FHWA-RD-88-009, 143 p.

Duan, Naihua, 1983, Smearing estimate—A nonparametric retransformation method: Journal of the American Statistical Association, v. 78, no. 383, p. 605–610.

Federal Highway Administration, 1998, Environmental flow charts: Washington, D.C., Department of Transportation, Federal Highway Administration, Office of Natural Environment, 45 p.

Federal Highway Administration, 2006, Water quality management of highway runoff—Participant workbook: Washington D.C., National Highway Institute, Federal Highway Administration, National Highway Institute Course Number 142047A, 296 p.

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., 1987, Sediment-transport curves: U.S. Geological Survey Open-File Report 87–218, 47 p.

Granato, G.E., 2006, Kendall-Theil Robust Line (KTRLine–version 1.0)—A Visual Basic program for calculating and graphing robust-nonparametric estimates of linear-regression coefficients between two continuous-variables: Techniques and Methods of the U.S. Geological Survey, chap. A7, book 4, 31 p.

Granato G.E., and Cazenas, P.A., 2007, Highway-runoff database (HRDB Version 1.0): A data warehouse and preprocessor for the stochastic empirical loading and dilution model: Washington, D.C., U.S. Department of Transportation, Federal Highway Administration, Office of Natural and Human Environment, Report FHWA-HEP-09-004, 57 p.

Granato, G.E., Bank, F.G., and Cazenas, P.A., 2003, Data quality objectives and criteria for basic information, acceptable uncertainty, and quality-assurance and quality-control documentation, *in* Granato, G.E., Zenone, C., and Cazenas, P.A. (eds.), National highway runoff water-quality data and methodology synthesis, Volume I—Technical issues for monitoring highway runoff and urban stormwater: Washington, D.C., U.S. Department of Transportation, Federal Highway Administration, FHWA-EP-03-054, p. 3–21.

Gray, J.R., Glysson, G.D., Turcios, L.M., Schwarz, G.E., 2000, Comparability of suspended-sediment concentration and total suspended solids data: U.S. Geological Survey Water-Resources Investigations Report 2000–4191, 14 p.

Griffith, J.A., Martinko, E.A., Whistler, J.L., and Price, K.P., 2002, Interrelationships among landscapes, NDVI, and stream water quality in the U.S. central plains: Ecological Applications, v. 12, no. 6, p. 1702–1718.

Haan, C.T., 1994, Statistical methods in hydrology: Ames, Iowa, Iowa State University Press, 378 p.

Heiskary, S., Wilson, C.B., and Larsen, D.P., 1987, Analysis of regional patterns in lake water quality—Using ecoregions for lake management in Minnesota: Lake and Reservoir Management, v. 3, p. 337–344.

Helsel, D.R., 2004, Nondetects and data analysis—Statistics for censored environmental data: Hoboken, N.J., John Wiley & Sons Publishing Inc., 268 p.

Helsel D.R., and Hirsch, R.M., 2002, Statistical methods in water resources—Hydrologic analysis and interpretation: Techniques of Water-Resources Investigations of the U.S. Geological Survey, chap. A3, book 4, 510 p.

Hem, J. D. 1992, Study and interpretation of the chemical characteristics of natural water, 3rd ed.: U.S. Geological Survey Water-Supply Paper 2254, 263 p.

Hirsch, R.M., Alexander, R.B., Smith, R.A., 1991, Selection of methods for the detection and estimation of trends in water quality: Water Resources Research, v. 27, no. 5, p. 803–813.

Hirsch, R.M., Helsel D.R., Gilroy, E.J., and Cohn, T.A., 1992, Chapter 17—Statistical analysis of hydrologic data, *in* Maidment D.R., (ed.), Handbook of hydrology: New York, McGraw-Hill Book Company, p. 17.1–17.55.

Hirsch, R.M., Slack, J.R., and Smith, R.A., 1982, Techniques of trend analysis for monthly water quality data: Water Resources Research, v. 18, p. 107–121.

Hosking, J.R.M., and Wallis, J.R., 1997, Regional frequency analysis—An approach based on L-moments: New York, N.Y., Cambridge University Press, 224 p.

House, W.A., and Warwick, M.S., 1998, Hysteresis of the solute concentration discharge relationship in rivers during storms: Water Research, v. 32, no. 8, p. 2279–2290.

Hren, J., Chaney, T.H., Norris, J.M., and Childress, C.J.O., 1987, Water-quality data-collection activities in Colorado and Ohio, Phase I—Inventory and evaluation of 1984 programs and costs: U.S. Geological Survey Water-Supply Paper 2295-A, 71 p.

Hughes, R.M., and Larsen, D.P., 1988, Ecoregions—An approach to surface water protection: Journal of the Water Pollution Control Federation, v. 60, no. 4, p. 486–493.

Intergovernmental Task Force on Monitoring Water Quality, 1995a, The strategy for improving water-quality monitoring in the United States: U.S. Geological Survey Final Report, 25 p.

Intergovernmental Task Force on Monitoring Water Quality, 1995b, The strategy for improving water-quality monitoring in the United States: U.S. Geological Survey Technical Appendixes, 117 p.

Jenerette, G.D., Lee, Jay, Waller, D.W., Carlson, R.E., 2002, Multivariate analysis of the ecoregion delineation for aquatic systems: Environmental Management, v. 29, no. 1, p. 67–75.

Jennings, M.E., Thomas, W.O. Jr., and Riggs, H.C., 1994, Nationwide summary of U.S. Geological Survey regional regression equations for estimating magnitude and frequency of floods for ungaged sites: U.S. Geological Survey Water-Resources Investigations Report 94-4002, 196 p.

Jones, B.E., 2003, Principles and practices for quality assurance and quality control *in* Granato, G.E., Zenone, C., and Cazenas, P.A. (eds.), National highway runoff water-quality data and methodology synthesis, Volume I—Technical issues for monitoring highway runoff and urban stormwater: Washington, D.C., U.S. Department of Transportation, Federal Highway Administration, FHWA-EP-03-054, p. 23–46.

Kirby, W.H., 1972, Computer-oriented Wilson-Hilferty transformation that preserves the first three moments and the lower bound of the Pearson type 3 distribution: Water Resources Research, v. 8, no. 5, p. 1251–1254.

Koch, R.W., and Smillie, G.M., 1986, Bias in hydrologic prediction using log-transformed regression models: Water Resources Bulletin, v. 22, no. 5, p. 717–723.

Langbein, W.B., and Schumm, S.A., 1958, Yield of sediment in relation to mean annual precipitation: Transactions of the American Geophysical Union, v. 39, no. 6, p. 1076–1084.

Lins, H.F., 1997, Regional streamflow regimes and hydroclimatology of the United States: Water Resources Research, v. 33, p. 1655–1667.

Litke, D.W., 1999, Review of phosphorus control measures in the United States and their effects on water quality: U.S. Geological Survey Water-Resources Investigations Report 99-4007, 38 p.

Lopes, T.J. and Dionne, S.G., 2003, A review of semivolatile and volatile organic compounds in highway runoff and urban stormwater, *in* Granato, G.E., Zenone, C., and Cazenas, P.A. (eds.), National highway runoff water-quality data and methodology synthesis, Volume I—Technical issues for monitoring highway runoff and urban stormwater: Washington, D.C., U.S. Department of Transportation, Federal Highway Administration, FHWA-EP-03-054, p. 235–303.

Marsalek, J. H., 1991, Pollutant loads in urban stormwater—Review of methods for planning-level estimates: Water Resources Bulletin, v. 27, no. 2, p. 283–291.

Marsalek, J. H. and Ng, Y.F., 1989, Evaluation of pollution loadings from urban nonpoint sources—Methodology and applications: Journal of Great Lakes Research, v. 15, no. 3, p. 444–451.

Mathey, S.B., (ed.), 1998, National Water Information System: U.S. Geological Survey Fact Sheet FS-027-98, 2 p.

Meade, R.H., and Parker, R.S., 1985, Sediment in rivers of the United States, *in* National water summary 1984—Hydrologic events, selected water-quality trends, and ground-water resources: U.S. Geological Survey Water Supply Paper 2275, p. 49–60.

Miller, C.R., 1951, Analysis of flow-duration, sediment-rating curve method of computing sediment yield: U. S. Bureau of Reclamation unnumbered report, 65 p.

Mueller, D.K. and Helsel, D.R., 1996, Nutrients in the Nation's waters too much of a good thing?: U.S. Geological Survey Circular 1136, 24 p.

Nash, D.B., 1994, Effective sediment-transporting discharge from magnitude-frequency analysis: Journal of Geology, v. 102, p. 79–95.

National Institute of Standards & Technology, 2003, NIST/SEMATECH e-handbook of statistical methods—Glossary: Accessed February 3, 2007 at *http://www.itl.nist.gov/div898/handbook/glossary.htm*

National Water-Quality Monitoring Council, 2006, National Water Quality Monitoring Council glossary of water-quality monitoring terms: Accessed February 3, 2007 at *http://acwi.gov/monitoring/glossary.html*

Norris, J.M., Hren, J., Myers, D., Chaney, T.H., and Childress, C.J.O., 1990, Water-quality data-collection activities in Colorado and Ohio, Phase III—Evaluation of existing data for use in assessing regional water-quality conditions and trends: U.S. Geological Survey Open-File Report 89-391, 63 p.

O'Connor, D.J., 1976, The concentration of dissolved solids and river flow: Water Resources Research, v. 12, no. 2, p. 279-294.

Omernik, J.M., 2004, Perspectives on the nature and definition of ecological regions—Environmental Management: Accessed on July 28, 2005, at *http://springerlink.metapress.com/openurl.asp?genre=article&id=doi:10.1007/s00267-003-5197-2*

Omernik J.M., 1995, Ecoregions—A spatial framework for environmental management, Chapter 5, *in* Davis, W.S. and Simon T.P. (eds.), Biological assessment and criteria—Tools for water resource planning and decision making: Boca Raton, FL, Lewis Publishers, p. 49–66.

Omernik, J.M., and Bailey, R.G., 1997, Distinguishing between watersheds and ecoregions: Journal of the American Water Resources Association, v. 33, no. 5, p. 935–949.

Omernik, J.M., Chapman, S. S., Lillie, R.A., Dumke, R.T., 2000, Ecoregions of Wisconsin: Transactions of the Wisconsin Academy of Sciences, Arts, and Letters, v. 88, p. 77–103.

Omernik J.M., and Powers, C.F., 1983, Total alkalinity of surface waters—A national map: Annals of the Association of American Geographers, v. 73, p. 133–136.

Rainwater, F.H., 1962, Stream composition of the conterminous United States: U.S. Geological Survey Hydrologic Investigations Atlas HA-61, 3 pls.

Riggs, H.C., 1968, Some statistical tools in hydrology: U.S. Geological Survey, Techniques of Water-Resources Investigations, book 4, chap. A1, 39 p.

Robertson, D.M., Saad, D.A., and Wieben, A.M., 2001, An alternative regionalization scheme for defining nutrient criteria for rivers and streams: U.S. Geological Survey Water-Resources Investigations Report 2001-4073, 57 p.

Rohm, C.M., Omernik, J.M., Woods, A.J., and Stoddard, J.L., 2002, Regional characteristics of nutrient concentrations in streams and their application to nutrient criteria development: Journal of the American Water Resources Association, v. 38, no. 1, p. 213–239.

Sen, P.K., 1968, Estimates of the regression coefficient based on Kendall's tau: Journal of the American Statistical Association, v. 63, p.1379–1389.

Sevin, A.F, 1987, Guidance for preparing and processing environmental and section 4(f) documents: Washington, D.C., Department of Transportation, Federal Highway Administration, Technical Advisory T6640.8A, 62 p.

Shirazi, M.A., Boersma, Larry, Johnson, C.B., and Haggerty, P.K., 2001, Predicting physical and chemical water properties from relationships with watershed soil characteristics: Journal of Environmental Quality, v. 30, p. 112–120.

Shirazi, M.A., Johnson, C.B., Omernik, J.M., White, D.A., Haggerty, P.K., and Griffith, G.E., 2003, Landscape and watershed processes—Quantitative soil descriptions for ecoregions of the Unites States: Journal of Environmental Quality, v. 32, no. 2, p. 550–561.

Shoemaker, L., Lahlou, M., Bryer, M., Kumar, D., and Kratt, K., 1997, Compendium of tools for watershed assessment and TMDL development: U.S. Environmental Protection Agency, Office of Water, EPA 841-B-97-006, 118 p.

Simon, Andrew, 2003, Suspended-sediment-transport rates at the 1.5-year recurrence interval for ecoregions of the United States—Transport conditions at the bankfull and effective discharge, *in* Renard, K.G., McElroy, S.A., Gburek, W.J., Canfield, H. E. and Scott, R.L., eds, 2003, First Interagency Conference on Research in the Watersheds, October 27-30, 2003: U.S. Department of Agriculture, Agricultural Research Service, p. 422–427.

Simon, Andrew, 1989, The discharge of sediment in channelized alluvial streams: Water Resource Bulletin, American Water Resources Association, v. 25, no. 6, p. 1177–1188.

Simon, Andrew, Dickerson, Wendy, and Heins, Amanda, 2004, Suspended sediment transport rates at the 1.5-year recurrence interval for ecoregions of the United States—Transport conditions at the bankfull and effective discharge: Geomorphology, v. 58, p. 243–262.

Sinnott, R.W., 1984, Virtues of the haversine: Sky and Telescope, v. 68, no. 2, p. 159.

Smith, R.A., Alexander, R.B., and Lanfear, K.J., 1993, Stream water quality in the conterminous United States—Status and trends of selected indicators during the 1980's, *in* Paulson, R.W., Chase E.B., Williams, J.S., and Moody, D.W. (eds.), National water summary 1990–91—Hydrologic events and stream water quality: U.S. Geological Survey Water-Supply Paper 2400, p. 111–140.

Smith, R.A., Alexander, R.B., and Schwarz, G.E., 2003a, Natural background concentrations of nutrients in streams and rivers of the conterminous United States: Environmental Science and Technology, v. 37, no. 4, p. 3039–3047.

Smith, R.A., Alexander, R.B., and Schwarz, G.E., 2003b, Natural background concentrations of nutrients in streams and rivers of the conterminous United States: Environmental Science and Technology, v. 37, no. 4, Supplemental Material, accessed on July 28, 2005, at *http://pubs3.acs.org/acs/journals/supporting_information.page?in_coden=esthag&in_volume=37&in_start_page=3039*

Smith, R.A., Alexander, R.B., Wolman, M.G., 1987, Analysis and interpretation of water-quality trends in major U.S. rivers, 1974–81: U.S. Geological Survey Water-Supply Paper 2307, 25 p.

Smith, R.A., Hirsch, R.M., and Slack, J.R., 1982, A study of trends in total phosphorus measurements at NASQAN stations: U.S. Geological Survey Water-Supply Paper 2190, 34 p.

Smith, R.A., Schwarz, G.E., Alexander, R.B., 1997, Regional interpretation of water-quality monitoring data: Water Resources Research, v. 33, no. 12, p. 2781–2798.

Stedinger, J.R., Vogel, R.M. and Foufoula-Georgiou, E., 1993, Frequency analysis of extreme events, Chapter 18, *in* Maidment, D.R. (ed.), Handbook of hydrology: McGraw-Hill Book Company, p. 18.1–18.66.

Syvitski J.P., Morehead, M.D., Bahr, D.B., and Mulder, Thierry, 2000, Estimating fluvial sediment transport: Water Resources Research, v. 36, no. 9, p. 2747–2760.

Tasker, G.D., and Granato, G.E., 2003, Statistical approaches to interpretation of local, regional, and national highway-runoff and urban-stormwater data, *in* Granato, G.E., Zenone, C., and Cazenas, P.A. (eds.), National highway runoff water-quality data and methodology synthesis, Volume I—Technical issues for monitoring highway runoff and urban stormwater: Washington, D.C., U.S. Department of Transportation, Federal Highway Administration, FHWA-EP-03-054, p. 419–479.

Theil, H., 1950, A rank-invariant method of linear and polynomial regression analysis, Part 3: Proceedings of Koninalijke Nederlandse Akademie van Weinenschatpen A, v. 53, p. 1397–1412.

Thomas, D.M., and Benson, M.A., 1970, Generalization of streamflow characteristics from drainage-basin characteristics: U.S. Geological Survey Water-Supply Paper 1975, 55 p.

Turcios, L.M., and Gray, J.R., 2001, U.S. Geological Survey sediment and ancillary data on the world wide web: Proceedings of the 7th Federal Interagency Sedimentation Conference, March 25–29, 2001, Reno, Nevada, Poster 31–36.

Urbonas, B.R. 1996, Determining urban stormwater best management practices (BMP) removal efficiencies—PA-ASCE Cooperative Agreement Task 2.2—Recommend the needed information/data tables and suggested database structure: Accessed on January 28, 2005, at *http://www.bmpdatabase.org/downloads.htm*

U.S. Environmental Protection Agency, 1986a, Development of data quality objectives: Washington, D.C., U.S. Environmental Protection Agency Quality Assurance Management Staff, 12 p.

U.S. Environmental Protection Agency, 1986b, Technical guidance manual for performing wasteload allocation—Book 6, Design conditions, Chapter I Stream design flow for steady-state modeling: Washington, D.C., U.S. Environmental Protection Agency Report EPA 440/4-86-014, 64 p.

U.S. Environmental Protection Agency, 1994, Guidance for the data quality objectives process, EPA QA/G-4, Final Report: Washington, D.C., U.S. Environmental Protection Agency, EPA/600/R-96/055, 72 p.

U.S. Environmental Protection Agency, 1991, Technical support document for water quality-based toxics control: Washington, D.C., U.S. Environmental Protection Agency, EPA/505/2-90-001, 336 p.

U.S. Environmental Protection Agency, 1994, Water quality standards handbook—2d ed.: Washington, D.C., U.S. Environmental Protection Agency, EPA 823-B-94-005a, 21 p.

U.S. Environmental Protection Agency, 1996, Guidance for data quality assessment, Practical methods for data analysis, EPA QA/G-9, Final Report: Washington, D.C., U.S. Environmental Protection Agency, EPA/600/R-96/084, 164 p.

U.S. Environmental Protection Agency, 1997, Guiding principles for Monte Carlo analysis: Washington, DC, U.S. Environmental Protection Agency, Risk Assessment Forum EPA/630/R-97/001, 39 p.

U.S. Environmental Protection Agency, 2002, National recommended water quality criteria–2002: Washington, D.C., U.S. Environmental Protection Agency, Office of Water, EPA-822-R-02-047, 43 p.

U.S. Environmental Protection Agency, 2000, Nutrient criteria technical guidance manual: Washington, D.C., U.S. Environmental Protection Agency, EPA-822-B-00-002.

U.S. Environmental Protection Agency, 2003a, Clean Water Act: Washington D.C., U.S. Environmental Protection Agency: Accessed on June 28, 2005, at *http://www.epa.gov/region5/water/cwa.htm*

U.S. Environmental Protection Agency, 2003b, Level III ecoregions of the continental United States: Washington, D.C., U.S. Environmental Protection Agency National Health and Environmental Effects Research Laboratory, 1 pl.

U.S. Environmental Protection Agency, 2004, Level III ecoregions: Corvallis, OR, U.S. Environmental Protection Agency, Western Ecology Division: Accessed on June 28, 2005, at *http://www.epa.gov/wed/pages/ecoregions/level_iii.htm*

U.S. Environmental Protection Agency, 2005a, Welcome to STORET, EPA's largest computerized environmental data system: Accessed on January 28, 2005, at *http://www.epa.gov/STORET/index.html*

U.S. Environmental Protection Agency, 2005b, Water quality criteria—Nutrients: Washington D.C., U.S. Environmental Protection Agency, Office of Water: Accessed on January 28, 2005, at *http://www.epa.gov/waterscience/standards/nutrient.html*

U.S. Environmental Protection Agency, 2005c, Water quality criteria—Suspended and bedded sediments: Washington D.C., U.S. Environmental Protection Agency, Office of Water: Accessed on January 28, 2005, at *http://www.epa.gov/waterscience/criteria/sediment/*

U.S. Environmental Protection Agency, 2006, Environmental Protection Agency's total maximum daily load (TMDL) program: Accessed on January 1, 2006, at *http://www.epa.gov/owow/tmdl/.*

U.S. Environmental Protection Agency, 2007a, Glossary of integrated risk information system terms: Accessed February 3, 2007 at *http://www.epa.gov/iris/gloss8.htm*

U.S. Environmental Protection Agency, 2007b, Mid-Atlantic States—Total maximum daily loads (TMDLs) primer glossary: Accessed February 3, 2007 at *http://www.epa.gov/ins/gloss8.htm*

U.S. Geological Survey, 1982, A U.S. Geological Survey data standard—Codes for the identification of hydrologic units in the United States and the Caribbean outlying areas: U.S. Geological Survey Circular 878-A, 115 p.

U.S. Geological Survey, 2002, NWIS Web—New site for the Nation's water data: U.S. Geological Survey Fact Sheet FS-128-02, 2 p.

U.S. Geological Survey, 2004a, NWIS Web data for the Nation: Accessed December 24, 2004 at *http://waterdata.usgs.gov.nwis/.*

U.S. Geological Survey, 2004b, Water basics glossary: Accessed February 3, 2007 at *http://capp.water.usgs.gov/GIP/h2o_gloss/*

U.S. Geological Survey, 2007, National Water-Quality Assessment (NAWQA) Program Glossary: Accessed February 3, 2007 at *http://water.usgs.gov/nawqa/glos.html*

U.S. National Atmospheric Deposition Program, 2004, Isopleth maps: Accessed December 24, 2004 at *http://nadp.sws.uiuc.edu/isopleths/.*

Vogel, R.M., Rudolph, B.E., and Hooper, R.P., 2005, Probabilistic behavior of water-quality loads: Journal of Environmental Engineering, v. 131, no. 7, p. 1081–1089.

Vogel, R.M., Stedinger, J.R., and Hooper, R.P., 2003, Discharge indices for water-quality loads: Water Resources Research, v. 39, no. 10, p. 1273–1281.

Vogel, R.M., Tsai, Y., and Limbrunner, J.F., 1998, The regional persistence and variability of annual streamflow in the United States: Water Resources Research, v. 34, no. 12, p. 3445–3459.

Warn, A.E., and Brew, J.S., 1980, Mass balance: Water Research, v. 14, p. 1427–1434.

Wilcox, R.R., 1998, A note on the Theil-Sen regression estimator when the regressor is random and the error term is heteroscedastic: Biometrical Journal, v. 40, p. 261–268.

Wilcox, R.R., 2001, Modern insights about Pearson's correlation and least-squares regression: International Journal of Selection and Assessment, v. 9, no. 1–2, p. 195–205.

Glossary

This glossary contains definitions of terms as they are used in this report. Some terms are defined in the glossary to help define other terms. Where possible, these definitions were adopted and adapted from published glossaries of the Federal Highway Administration (Jones, 2003; Federal Highway Administration, 2006), the National Institute of Standards and Technology (2003), the National Water-Quality Monitoring Council (2006), U.S. Environmental Protection Agency (1991; 1994; 1997; 2006; 2007), and the U.S. Geological Survey (2004; 2007).

A

Accuracy The degree to which a measured value agrees with the true value of the measured property (Jones, 2003). The difference between accuracy and precision is that measures of accuracy degrade as bias increases, whereas measures of precision are not affected by bias (National Institute of Standards and Technology, 2003).

Acute A stimulus severe enough to rapidly induce an effect; in aquatic toxicity tests, an effect observed in 96 hours or less typically is considered acute. When referring to aquatic toxicology or human health, an acute effect is not always measured in terms of lethality (U.S. Environmental Protection Agency, 1991).

Ambient concentration Concentrations of a substance in a particular environment that are characteristic of the surrounding environment and may include the spectrum anthropogenic influence from natural background concentrations to concentrations affected by point and nonpoint source pollution of air, soil, and water.

Anthropogenic Characteristic of or occurring because of, or influenced by, human activity (U.S. Geological Survey, 2004b; 2007).

Aquatic ecosystem The stream channel, lake or estuary bed, water, and (or) biotic communities and the habitat features that occur therein (National Water-Quality Monitoring Council, 2006).

Aquatic guidelines Specific levels of water quality which, if reached, may adversely affect aquatic life. These are nonenforceable guidelines issued by a governmental agency or other institution (U.S. Geological Survey, 2004b; 2007).

Aquatic-life criteria Water-quality guidelines for protection of aquatic life. Often refers to U.S. Environmental Protection Agency water-quality criteria for protection of aquatic organisms. See also Water-quality guidelines, Water-quality criteria, and Freshwater chronic criteria. (U.S. Geological Survey, 2004b; 2007).

B

Background concentration A (natural) concentration of a substance in a particular environment that is indicative of minimal influence by anthropogenic sources of pollution to air, soil, and water (U.S. Geological Survey, 2004b; 2007).

Base flow The sustained low flow of a stream, usually ground-water inflow to the stream channel (U.S. Geological Survey, 2004b; 2007).

Bedload Sediment that moves on or near the streambed and is in almost continuous contact with the bed (U.S. Geological Survey, 2007).

Bed sediment The material that temporarily is stationary in the bottom of a stream or other watercourse (U.S. Geological Survey, 2007).

Best Management Practices (BMPs) Methods, measures, or practices that are determined to be reasonable and cost-effective means for a landowner to meet certain, generally nonpoint source, pollution control needs. BMPs include structural and nonstructural controls and operation and maintenance procedures (U.S. Environmental Protection Agency, 2007b).

Bias The extent to which a measured value differs from the true value of the measured property (Jones, 2003).

Blank A synthetic sample that is (initially) free of the analyte(s) of interest (Jones, 2003).

C

Chronic A stimulus that can cause or contribute to observations of reduced growth, reduced reproduction, mutation, disease, and potentially lethality if exposure to the stimulus lingers or continues for a relatively long period of time, often one-tenth of the life span or more. Chronic should be considered a relative term depending on the life span of an organism (U.S. Environmental Protection Agency, 1991).

Clean Water Act (CWA) The Clean Water Act (formerly referred to as the Federal Water Pollution Control Act or Federal Water Pollution Control Act Amendments of 1972), Public Law 92-500, as amended by Public Law 96-483 and Public Law 97-117, 33 U.S.C. 1251 et seq. The Clean Water Act (CWA) contains a number of provisions to restore and maintain the quality of the nation's water resources. One of these provisions is section 303(d), which establishes the total maximum daily loads (TMDL) program (U.S. Environmental Protection Agency, 2007b).

Comparability The degree to which two pieces or sets of data represent environmental conditions in the same way, so that the difference between the data or sets of data accurately represents the change in environmental conditions over time or space (Jones, 2003). The characteristics that allow data produced by multiple methods to meet or exceed the data-quality objectives of primary or secondary data users. These characteristics need to be defined to establish comparability and would likely include data-quality objectives, bias, precision, data collection, handling and processing methods, and other information (National Water-Quality Monitoring Council, 2006)

Concentration The amount or mass of a substance present in a given volume or mass of sample. Usually expressed as milligrams or microgram per liter (water sample) or as milligrams or micrograms per kilogram (sediment or tissue sample) (U.S. Geological Survey, 2004b; 2007).

Confidence interval A statement that an unknown parameter is between two values with a certain probability. Based on the observation of a certain set of data, the range of plausible values of an unknown parameter that are consistent with observing that data. For example, if one says the 95-percent confidence interval for a variable is 1.1 to 10.3, then this is equivalent to saying that based on the data observed, there is a 95-percent chance that the variable is between 1.1 and 10.3 (National Institute of Standards and Technology, 2003).

Constituent A chemical, physical, or biological substance in water, sediment, or biota that can be measured by an analytical method (U.S. Geological Survey, 2004b; 2007).

Contaminant A material added by humans or natural activities that may, in sufficient concentrations, render the environment unacceptable for biota. The mere presence of these materials is not necessarily harmful (National Water-Quality Monitoring Council, 2006).

Contamination Degradation of water quality compared to original or natural conditions due to human activities or natural processes (U.S. Geological Survey, 2004b; 2007; National Water-Quality Monitoring Council, 2006).

Conterminous United States The 48 U.S. States enclosed within a common boundary in North America. Not including Alaska and Hawaii, which do not have the equivalent information and data for synoptic storm events, streamflow, water quality, and USEPA level III nutrient ecoregions.

Correlation, Correlation Analysis An investigation of the measure of statistical association among random variables based on samples. Widely used measures include the linear correlation coefficient (also called the product-moment correlation coefficient or Pearson's correlation coefficient), and such non-parametric measures as Spearman rank-order correlation coefficient, and Kendall's tau. When the data are nonlinear, non-parametric correlation is generally considered to be more robust than linear correlation (U.S. Environmental Protection Agency, 1997).

Coefficient of variation (CV) A standard statistical measure of the relative variation of a distribution or set of data, defined as the standard deviation divided by the mean (U.S. Environmental Protection Agency, 1991).

Context-sensitive solutions (CSS) A collaborative, interdisciplinary approach that involves all stakeholders to develop a transportation facility that fits its physical setting and preserves scenic, aesthetic, historic, and environmental resources, while maintaining safety and mobility (Federal Highway Administration, 2006).

Criteria continuous concentrations (CCC) The USEPA national water quality criteria recommendation for the highest instream concentration of a contaminant or an effluent to which organisms can be exposed for a brief period of time without causing a chronic effect (U.S. Environmental Protection Agency, 1991).

Criteria maximum concentrations (CMC) The USEPA national water quality criteria recommendation for the highest instream concentration of a contaminant or an effluent to which organisms can be exposed for a brief period of time without causing an acute effect (U.S. Environmental Protection Agency, 1991).

Criterion A standard rule or test on which a judgment or decision can be based (U.S. Geological Survey, 2004b; 2007). Criteria are elements of state water-quality standards, expressed as constituent concentrations, levels, or narrative statements, representing a quality of water that supports a particular use (40 CFR 131.3.). When criteria are met, water quality will generally protect the designated use (U.S. Environmental Protection Agency, 1994).

Cubic foot per second (ft^3/s, or cfs) The rate of water discharge representing a volume of 1 cubic foot passing a given point during 1 second, equivalent to approximately 7.48 gallons per second or 448.8 gallons per minute or 0.02832 cubic meter per second (U.S. Geological Survey, 2004b; 2007).

Cubic feet per second per square mile ($ft^3/s/mi$, or cfsm) The rate of water discharge in cubic feet per second normalized to (divided by) the upstream drainage area usually used for comparison between sites of different drainage areas and estimation of discharge at hydrologically similar ungaged sites.

D

Data-quality objectives Qualitative and quantitative statements that clarify the goals of the study, define the appropriate data to be collected, determine the most appropriate conditions from which to collect the data, and specify tolerance limits on decision errors that will be used to establish the quality and quantity of data needed to support the decision (Jones, 2003).

Decisionmakers Planners, regulators, highway engineers, and other stakeholders who may have input on planning, designing, building, and maintaining transportation facilities.

Dependent variable The value predicted (commonly designated as *Y*) from a mathematical model from an independent (predictor) variable (commonly designated as *X*). See independent variable.

Detect To determine the presence of a compound (U.S. Geological Survey, 2004b).

Detection limit see Limit of detection.

Deterministic The property of being perfectly repeatable, and without experimental or observational error, usually achievable only in computer experiments (National Institute of Standards and Technology, 2003). A process or variable that can be predicted with a mathematical model perfectly without random variation or error (Chow, 1971). For comparison see probabilistic and stochastic.

Digestion method Standard chemical methods that prescribe reagents, concentrations of reagents, temperature, and contact time designed to solubilize sediment associated trace elements to facilitate analysis of constituent concentrations. These standard methods may solubilize different percentages of the constituent of interest in samples with differing water chemistry, sediment chemistry, and sediment concentrations (Breault and Granato, 2003).

Dilution A reduction in concentration caused by mixing waters with different concentrations. For example, a volume of water with a higher concentration may be diluted by addition of a volume of water with a lower concentration of a given constituent.

Discharge The volume of fluid passing a point per unit of time, commonly expressed million gallons per day, gallons per minute, or cubic feet per second (U.S. Geological Survey, 2004b; 2007).

Dissolved constituent Operationally defined as a constituent that passes through a 0.45-micrometer filter (U.S. Geological Survey, 2004b; 2007).

Dissolved solids Amount of minerals, such as salt, that are dissolved in water; amount of dissolved solids is an indicator of salinity or hardness (U.S. Geological Survey, 2004b; 2007).

Distribution A representation of the frequency of occurrence of values of a variable, especially of a response variable (National Institute of Standards and Technology, 2003).

Drainage area The area (measured in a horizontal plane) that is enclosed by a drainage divide, which encompasses a drainage basin upstream of a specified location along a stream (U.S. Geological Survey, 2004b; 2007).

Drainage basin The portion of the surface of the Earth that contributes water to a stream at a specified location along a stream through overland runoff, including tributaries and impoundments (U.S. Geological Survey, 2004b; 2007).

Drainage divide Boundary between adjoining drainage basins for a specified location along a stream, commonly following topographic high points (U.S. Geological Survey, 2004b; 2007).

E

Ecoregion Regions of ecological similarity defined and delineated by a semi-quantitative or qualitative process that uses information about geology, physiography, vegetation, climate, soils, land use, wildlife, and hydrology and other ecologically relevant variables to identify a homogeneous area. Regions of ecological similarity help define the potential designated use classifications of specific water bodies (Omernik and others, 2000; Omernik, 2004; National Water-Quality Monitoring Council, 2006).

Ecosystem The interacting populations of plants, animals, and microorganisms occupying an area, plus their physical and chemical environment (National Water-Quality Monitoring Council, 2006; U.S. Geological Survey, 2007).

Effluent Outflow from a particular source, such as a stream that flows from a lake or liquid waste that flows from a factory or sewage-treatment plant (U.S. Geological Survey, 2004b; 2007).

Ephemeral stream A stream or part of a stream that flows only in direct response to precipitation or snowmelt. Its channel is above the water table at all times (U.S. Geological Survey, 2007).

Erosion The process whereby materials of the Earth’s crust are loosened, dissolved, or worn away and simultaneously moved from one place to another (U.S. Geological Survey, 2004b; 2007).

Event mean concentrations (EMCs) The total water-quality constituent load from a storm event divided by the total volume of runoff from a storm event.

F

Far outlier Data that is beyond two orders of magnitude higher or lower than the nearest point and being of questionable physical veracity.

Fluvial Pertaining to a river or stream (U.S. Geological Survey, 2004b; 2007).

Fluvial deposit A sedimentary deposit consisting of material transported by suspension or laid down by a river or stream (U.S. Geological Survey, 2007)

Freshwater Water that contains less than 1,000 milligrams per liter of dissolved solids (U.S. Geological Survey, 2004b).

G

Gaging station A particular site on a stream, canal, lake, or reservoir where systematic observations of hydrologic data are obtained (U.S. Geological Survey, 2004b; 2007).

Geographic information systems (GIS) A computerized system for combining, displaying, and analyzing geographic data. GIS produces maps for environmental planning and management by integrating physical and biological information (soils, vegetation, hydrology, living resources, and so forth) and cultural information (population, political boundaries, roads, bank and shoreline development, and so forth (U.S. Environmental Protection Agency, 2007b).

Goodness-of-fit As a general concept, the opposite of lack of fit. Any measure of how close a probability model reproduces the frequencies of an observed distribution. A measure, such as R-squared, of how close a statistical model predicts observed values. For comparison, see lack of fit (National Institute of Standards and Technology, 2003).

H

Hardness A property of water that causes the formation of an insoluble residue when the water is used with soap and a scale in vessels in which water has been allowed to evaporate. It is due primarily to the presence of ions of calcium and magnesium. Generally expressed as milligrams per liter as calcium carbonate ($CaCO_3$). General hardness categories are soft water (0–60 milligrams per liter as calcium carbonate), moderately hard (61–120 milligrams per liter as calcium carbonate), hard water (121–180 milligrams per liter as calcium carbonate), and very hard water (more than 180 milligrams per liter as calcium carbonate) (U.S. Geological Survey, 2004b). Hardness is an important trace-metal toxicity indicator (U.S. Environmental Protection Agency, 2002).

Heteroscedasticity Non-constant variance in the residuals of a regression model.

High-leverage outlier A data outlier (see far outlier) that has a substantial effect on calculation of coefficients of a parametric linear-regression equation.

Hydrograph Graph showing variation of water elevation, velocity, streamflow, or other property of water with respect to time (U.S. Geological Survey, 2004b; 2007b).

Hydrologic unit A geographic area representing part or all of a surface drainage basin or distinct hydrologic feature as delineated by the U. S. Geological Survey on State Hydrologic Unit Maps. Each hydrologic unit is assigned a hierarchical hydrologic unit code consisting of 2 digits for each successively smaller drainage basin unit (U.S. Geological Survey, 2004b).

Hydrology The science that deals with water as it occurs in the atmosphere, on the surface of the ground, and underground (U.S. Geological Survey, 2004b).

Hysteresis A physical effect in which a relation between variables depends on prior condition. For example, when sediment concentrations are higher for a given water discharge on the rising limb of the hydrograph than on the falling limb because erodible sediment is accumulated in a basin between storm events.

I

Impact A change in the chemical, physical, or biological quality or condition of a water body caused by external sources (U.S. Environmental Protection Agency, 2007b).

Independent variable The predictor value (commonly designated as *X*) input to a mathematical model to estimate the value of a dependent or predicted variable (commonly designated as *Y*) variable. See dependent variable.

Instantaneous discharge The volume of water that passes a point at a particular instant of time (U.S. Geological Survey, 2004b; 2007). In this report, the discharge measurement associated with the collection of a concurrent water-quality sample.

Intermittent stream A stream that flows when it receives water from rainfall runoff or springs, or from some surface source such as melting snow (U.S. Geological Survey, 2004b; 2007). Can flow from ground-water discharge (see base flow), at some times, but will go dry under relatively normal (non-drought) conditions.

K

Kendall-Theil robust line A nonparametric regression method in which the slope is calculated as the median of all pairwise slopes and the intercept is calculated so that the line passes through the medians of the independent (predictor) and dependent (predicted) variables (Granato, 2006).

L

Lack of fit A property of a model with respect to a set of observations. Lack of fit refers to the degree to which the model does not predict or fit the observations. Lack of fit can be due to experimental error, uncertainty in the process obtaining the observations, random (or unexplained) variation in the data, or a defect in the model (National Institute of Standards and Technology, 2003).

Limit of detection The lowest concentration of an analyte that can reliably be distinguished from a blank by a given analytical method or measurement. Generally computed as three times the standard deviation of the blank (Jones, 2003). The concentration below which a particular analytical method cannot determine, with a high degree of certainty, a concentration (U.S. Geological Survey, 2004b).

Limit of quantitation The lowest concentration that can be measured quantitatively by a given analytical method. Generally computed as 10 times the standard deviation of the blank (Jones, 2003).

Load General term that refers to a material or constituent in solution, in suspension, or in transport; usually expressed in terms of mass or volume (U.S. Geological Survey, 2004b; 2007).

Logarithmic transformation Use of the logarithms of data to improve the statistical properties of data for analysis or model specification. See transformation.

Lognormal distribution Data distribution in which the logarithms of the data form a normal distribution.

M

Major ions Constituents commonly present in concentrations exceeding 1.0 milligram per liter. Dissolved cations generally are calcium, magnesium, sodium, and potassium; the major anions are sulfate, chloride, fluoride, nitrate, and those contributing to alkalinity, most generally assumed to be bicarbonate and carbonate (U.S. Geological Survey, 2004b; 2007).

Maximum reporting limits The largest measured concentration of a constituent that may be reliably reported using a given analytical method.

Mean The average of a set of observations, unless otherwise specified (U.S. Geological Survey, 2007).

Median The middle or central value in a distribution of data ranked in order of magnitude. The median is also known as the 50th percentile (U.S. Geological Survey, 2007).

Measurement error The variability observed that can be attributed to the metrology or measurement system. Measurement error can be decomposed further into miscalibration, sensitivity, repeatability, and reproducibility (National Institute of Standards and Technology, 2003).

Metadata Information that describes the content, quality, condition, and other characteristics of data (National Water-Quality Monitoring Council, 2006).

Method error The part of the measurement error attributable to the details of the measurement process (National Institute of Standards and Technology, 2003).

Micrograms per liter (µg/L) A unit expressing the concentration of constituents in solution as weight (micrograms) of solute per unit volume (liter) of water; equivalent to one part per billion in most streamwater and ground water. One thousand micrograms per liter equals 1 mg/L (U.S. Geological Survey, 2007).

Milligrams per liter (mg/L) A unit expressing the concentration of chemical constituents in solution as weight (milligrams) of solute per unit volume (liter) of water; equivalent to one part per million in most streamwater and ground water. One thousand milligrams per liter equals 1 g/L. (U.S. Geological Survey, 2007).

Minimum reporting level (MRL) The smallest measured concentration of a constituent that may be reliably reported using a given analytical method. In many cases, the MRL is used when documentation for the method detection limit is not available (U.S. Geological Survey, 2004b; 2007).

Mitigation Actions, taken to avoid, reduce, or compensate for the effects of human-induced environmental damage (U.S. Geological Survey, 2004b). The Council on Environmental Quality defines mitigation as: avoiding the impact altogether by not taking a certain action or parts of an action; minimizing impacts by limiting the degree or magnitude of the action and its implementation; rectifying the impact by repairing, rehabilitating, or restoring the affected environment; reducing or eliminating the impact over time by preservation and maintenance operations during the life of the action; compensating for the impact by replacing or providing substitute resources or environments (40 CFR 1508.20; Patricia Cazenas, Federal Highway Administration, Office of Natural and Human Environment, written commun., 2007).

Model A mathematical statement of the relation(s) among variables or a representation of a physical system or process with a mathematical equation or equations (National Institute of Standards and Technology, 2003).

Monte Carlo method A computer experimental method that uses random numbers in order to estimate distributions of simulator outputs (National Institute of Standards and Technology, 2003). A stochastic modeling technique that involves the random selection of sets of input data for use in repetitive model runs in order to predict the probability distributions of receiving water quality concentrations (U.S. Environmental Protection Agency, 1991).

N

National Pollutant Discharge Elimination System (NPDES) The national program for issuing, modifying, revoking and reissuing, terminating, monitoring, and enforcing permits, and imposing and enforcing pretreatment requirements, under Sections 307, 402, 318, and 405 of the Clean Water Act. Facilities subjected to NPDES permitting regulations include operations such as municipal wastewater-treatment plants and industrial waste-treatment facilities (U.S. Environmental Protection Agency, 2007b).

Natural background levels Natural background levels represent the chemical, physical, and biological conditions that result from natural changes to the Earth's surface, such as weathering or decay (U.S. Environmental Protection Agency, 2007b).

Nitrate A negatively charged ion consisting of nitrogen and oxygen (NO_3^-). Nitrate is a plant nutrient and is very mobile in soils (U.S. Geological Survey, 2007).

Nonpoint source Pollution that is not released through pipes but rather originates from multiple sources over a relatively large area. Nonpoint sources can be divided into source activities related to either land or water use including failing septic tanks, improper animal-keeping practices, forestry practices, and urban and rural runoff (U.S. Environmental Protection Agency, 2007b). A contributory factor to water pollution that cannot be traced to a specific spot; for example, pollution that results from water runoff from urban areas, construction sites, agricultural and silvicultural operations, and so forth (National Water-Quality Monitoring Council, 2006).

Normal distribution A symmetric distribution with one high point or mode, sometimes also called the bell curve. The average is one of many statistical calculations that, even for only a moderate amount of data, tend to have a distribution that resembles the normal curve. There are four important properties of the normal distribution: (1) it is symmetric; (2) within plus or minus one standard deviation about 68 percent; (3) within plus or minus two standard deviations, 95 percent; and (4) within plus or minus three standard deviations, 99.7 percent of the distribution is enclosed (National Institute of Standards and Technology, 2003).

Normalized water discharge Water discharge values divided by the upstream drainage area of the monitoring site to allow comparison data from different measured sites or application of regional values to ungaged sites by use of drainage area ratios. Normalized water discharge is expressed as a rate per unit area, for example in cubic feet per second per square mile.

Nutrient Element or compound essential for animal and plant growth. Common nutrients in fertilizer include nitrogen, phosphorus, and potassium (U.S. Geological Survey, 2007).

O

Organic Molecules containing multiple carbon atoms, but possibly also containing hydrogen, oxygen, chlorine, nitrogen, and other elements. Examples are semivolatile organic coumpounds and volatile organic compounds (U.S. Geological Survey, 2004b).

Order-of-magnitude An increase or decrease in values that is a factor of ten times or one tenth of a given value, respectively; corresponding to one log scale on a base-10 logarithmic axis.

Outlier An observation whose value is so extreme that it appears not to be consistent with the rest of the dataset. Outliers indicate that assignable or special causes are present. The deletion of a particular outlier from a data analysis is easiest to justify when such an unusual cause has been identified (National Institute of Standards and Technology, 2003).

Overland flow The flow of rainwater or snowmelt over the land surface toward stream channels (U.S. Geological Survey, 2004b).

P

Part per million (ppm) Unit of concentration equal to one milligram per kilogram or one milligram per liter (U.S. Geological Survey, 2007).

Perennial stream A stream that normally has water in its channel at all times (U.S. Geological Survey, 2007). See intermittent and ephemeral stream.

pH The logarithm of the reciprocal of the hydrogen ion concentration (activity) of a solution; a measure of the acidity (pH less than 7) or alkalinity (pH greater than 7) of a solution; a pH of 7 is neutral (U.S. Geological Survey, 2007).

Phosphorus A nutrient essential for growth that can play a key role in stimulating aquatic growth in lakes and streams (U.S. Geological Survey, 2007).

Physiography A description of the surface features of the Earth, with an emphasis on the origin of landforms (U.S. Geological Survey, 2004b).

Planning-level estimates The results of analyses used to evaluate broad policy measures, which are recognized to include substantial uncertainties (commonly orders of magnitude) in all aspects of the decision process.

Point source A source of pollutants at a discrete location such as a discharge pipe, drainage ditch, tunnel, well, concentrated livestock operation, or floating craft (U.S. Geological Survey, 2007). Pollutant loads discharged at a specific location from pipes, outfalls, and conveyance channels from either municipal wastewater-treatment plants or industrial waste-treatment facilities. Point sources can also include pollutant loads contributed by tributaries to the main receiving water stream (U.S. Environmental Protection Agency, 2007b).

Pollutant A contaminant that is discharged to a water body, resulting in the impairment of that water body. Types of pollutants include dredged spoil, solid waste, incinerator residue, sewage, garbage, sewage sludge, munitions, chemical wastes, biological materials, radioactive materials, heat, wrecked or discarded equipment, rock, sand, cellar dirt, and industrial, municipal, and agricultural waste discharged into water (U.S. Environmental Protection Agency, 2007b). Any substance that, when present in a hydrologic system at sufficient concentration, degrades water quality in ways that are or could become harmful to human and/or ecological health or that impair the use of water for recreation, agriculture, industry, commerce, or domestic purposes (U.S. Geological Survey, 2004b; 2007).

Population The entire set of potential observations about whose properties we would like to learn. As opposed to sample. (National Institute of Standards and Technology, 2003).

Precipitation Any or all forms of water particles that fall from the atmosphere, such as rain, snow, hail, and sleet (U.S. Geological Survey, 2007).

Precision The variability of a measurement process around its average value. Precision is usually distinguished from accuracy, the variability of a measurement process around the true value. Precision, in turn, can be decomposed further into short-term variation or repeatability, and long-term variation, or reproducibility (National Institute of Standards and Technology, 2003).

Probabilistic The uncertain nature of a completely random variable whose specific value cannot be predicted exactly except in terms of the chance of occurrence (Chow, 1971). For comparison see deterministic and stochastic.

Probability A number expressing the likelihood of occurrence of a specific event, such as the ratio of the number of outcomes that will produce a given event to the total number of possible outcomes (U.S. Environmental Protection Agency, 1991).

Probability distribution A mathematical representation of the probabilities that a given variable will have various values (U.S. Environmental Protection Agency, 1991).

Probability plot A plot designed to assess whether an observed distribution has a shape consistent with a theoretical distribution, especially with the normal distribution. The values observed are plotted against the expected order statistics from the theoretical distribution. When a straight line is apparent, the observed and theoretical distributions are said to have the same shape. Probability plots are especially good when the observed distribution consists of many observations, and useful for comparing at most only a few groups (National Institute of Standards and Technology, 2003).

Q

Quality assurance Evaluation of quality-control data to allow quantitative determination of the quality of chemical data collected during a study. Techniques used to collect, process, and analyze water samples are evaluated. All of the planned and systematic activities implemented within the quality system, and demonstrated as needed, to provide adequate confidence that an entity will fulfill requirements for quality (Jones, 2003, U.S. Geological Survey, 2004b; 2007).

Quality control Operational techniques and activities that are used to fulfill requirements for quality (Jones, 2003).

Qualified values Reported water-quality or discharge data that include zeros, less than symbols, greater than symbols, estimated, and altered values denoting additional uncertainties in such data.

R

Random variable A random variable is a quantity which can take on any number of values but whose exact value cannot be known before a direct observation is made. For example, the outcome of the toss of a pair of dice is a random variable, as is the height or weight of a person selected at random from the New York City phone book (U.S. Environmental Protection Agency, 1997).

Receiving waters Streams, ponds, lakes, wetlands or other water bodies that receive point or nonpoint discharges of runoff. See Waters of the United States.

Reference value/conditions A single measurement or set of selected measurements of unimpaired water bodies characteristic of an ecoregion and (or) habitat. The chemical, physical, or biological quality or condition that is exhibited at either a single site or an aggregation of sites that represent the least impacted or reasonably attainable condition at the least impacted reference sites (U.S. Environmental Protection Agency, 2007a).

Reference site A sampling site selected for its relatively undisturbed conditions (U.S. Geological Survey, 2004b).

Regional analysis An analysis done on the assumption that groups of data from hydrologically similar sites defined by regional characteristics will be representative of conditions at unmonitored sites (or sites with limited data) within the identified region.

Regression A statistical method for deriving a mathematical model (usually a linear or exponential equation) to predict the most probable value of a dependent variable from a known independent or predictor variable with a known (or estimated) level of uncertainty or variation from the model-defined value.

Repeatability The component of measurement precision that is the variability characteristic of using the same or similar methods, instruments, operators, experimental setup, and ambient environmental conditions (National Institute of Standards and Technology, 2003). Documentation of project methods and other information that allows an independent researcher to achieve the same or similar results.

Representativeness The extent to which a sample represents the population from which it is withdrawn (U.S. Environmental Protection Agency, 1997; Jones, 2003).

Reproducibility The total measurement precision, especially including the components of variability that occur in the long term, and occurring from one measurement instrument to another, one laboratory to another, and so forth (National Institute of Standards and Technology, 2003).

Residual The difference between the actual value observed and the prediction or fitted value derived from a model. Residuals give information both about the model's lack of fit, and also about experimental error of the measurement process (National Institute of Standards and Technology, 2003).

Robust methods Methods of data analysis that are robust are not strongly affected by extreme changes to small portions of the data; their answers do not change very much from the presence of outliers. A classic example of a robust method is the median (National Institute of Standards and Technology, 2003).

Runoff Excess rainwater or snowmelt that is transported to streams by overland flow, tile drains, or ground water (U.S. Geological Survey, 2007). That part of precipitation, snowmelt, or irrigation water that runs off the land into streams or other surface water. It can carry pollutants from the air and land into receiving waters (U.S. Environmental Protection Agency, 1997).

S

Sample The set of observational units (wafers, people, etc.) whose properties our study is to observe. When we select a sample by scientific randomization, we are more easily able to generalize our conclusions to the population of interest. For a given characteristic, the collection of measurements that are actually observed. As opposed to population (National Institute of Standards and Technology, 2003).

Sample size The number of observations in, or planned to be in, a study or other investigation. Key considerations in selecting a particular sample size are value associated with any particular level of precision, the costs of obtaining observations, and available resources. Some generic rules of thumb on sample sizes are 16 observations to estimate the center of a distribution by its average; 20 (paired) observations to estimate the correlation between two measurements; 32 observations per group, to estimate average difference between two groups; 50 observations to estimate the standard deviation of a distribution (National Institute of Standards and Technology, 2003).

Sampling distribution The distribution of a summary quantity or statistic (National Institute of Standards and Technology, 2003).

Scatterplot A graph of a pair of variables that plots the first variable along the x-axis and the second variable along the y-axis. In a scatterplot, the points of successive pairs are not connected (National Institute of Standards and Technology, 2003).

Sediment Particles, derived from rocks or biological materials that have been transported by a fluid or other natural process, suspended or settled in water (U.S. Geological Survey, 2007).

Semivolatile organic compound (SVOC) Operationally defined as a group of synthetic organic compounds that are solvent-extractable and can be determined by gas chromatography/mass spectrometry. SVOCs include phenols, phthalates, and polycyclic aromatic hydrocarbons (PAHs) (U.S. Geological Survey, 2007).

Skewness Numerical measure of the lack of symmetry of an asymmetrical frequency distribution (U.S. Geological Survey, 2007).

Solute See Solution (U.S. Geological Survey, 2007).

Solution Formed when a solid, gas, or another liquid in contact with a liquid becomes dispersed homogeneously throughout the liquid. The substance, called a solute, is said to dissolve. The liquid is called the solvent (U.S. Geological Survey, 2007).

Solvent See Solution (U.S. Geological Survey, 2007).

Sorption General term for the interaction (binding or association) of a solute ion or molecule with a solid (U.S. Geological Survey, 2004b).

Standard deviation A measure of spread or dispersion of a distribution. It estimates the square root of the average squared deviation from the distribution average, sometimes called the root-mean-square. Among all measures of dispersion, the standard deviation is the most efficient for normally distributed data. Also, unlike the range, it converges to a single value as more data from the distribution is gathered (National Institute of Standards and Technology, 2003). It is the square root of the variance, which is calculated as the sum of the squares of the deviations from the arithmetic mean, divided by the number of values in the series minus 1 (U.S. Geological Survey, 2004b).

Standard error The standard deviation for a statistic's sampling distribution. Because many data sets have sampling distributions that are approximately normal (or lognormal), plus or minus 2 standard errors is usually an approximate 95-percent confidence interval (National Institute of Standards and Technology, 2003).

Statistic A value calculated from (an therefore somewhat characteristic of) sample data (National Institute of Standards and Technology, 2003).

Statistical significance The probability that a result is not likely to be due to chance alone. By convention, a difference between two groups is usually considered statistically significant if chance could explain it only 5 percent of the time or less. Study design considerations may influence the a priori choice of a different level of statistical significance (U.S. Environmental Protection Agency, 1997).

Statistics A branch of mathematics dealing with the collection, analysis, interpretation, and presentation of masses of numerical data (U.S. Geological Survey, 2007).

Stochastic A random process with a deterministic component so that a family of random variables depend on a parameter or parameters that belong to an indexing set causing the variables to have a different probability distribution for each value of a predictor or explanatory variable (Chow, 1971). Deterministic and probabilistic processes are special cases of a stochastic process. If the probability of the random variable approximates one for a given condition the stochastic process reduces to the deterministic case. If the probability of an individual value of the random variable is independent of the predictor or explanatory variable the process is purely probabilistic (with respect to the predictor variable) (Chow, 1971). For comparison see deterministic and probabilistic.

STORET USEPA's computerized water-quality database that includes physical, chemical, and biological data measured in water bodies throughout the United States (U.S. Environmental Protection Agency, 1991).

Streamflow A type of channel flow, applied to that part of surface runoff in a stream whether or not it is affected by diversion or regulation (U.S. Geological Survey, 2007). The discharge of water in a natural channel (U.S. Geological Survey, 2004b).

Sublethal A stimulus below the level that causes death (U.S. Environmental Protection Agency, 1991).

Surface runoff Runoff that travels over the land surface to the nearest stream channel (U.S. Geological Survey, 2004b).

Surface water An open body of water, such as a lake, river, or stream (U.S. Geological Survey, 2007).

Suspended The amount (concentration) of undissolved material in a water-sediment mixture. It is associated with the material retained on a 0.45-micrometer filter (U.S. Geological Survey, 2007).

Suspended sediment Particles of rock, sand, soil, and organic detritus carried in suspension in the water column, in contrast to sediment that moves on or near the streambed (U.S. Geological Survey, 2007).

Suspended sediment concentration The concentration of suspended sediment in the sampled zone (commonly from the water surface to a point approximately 0.3 foot above the bed) expressed as milligrams of dry sediment per liter of water-sediment mixture (mg/L) (U.S. Geological Survey, 2007).

Suspended solids Different from suspended sediment only in the way that the sample is collected and analyzed (U.S. Geological Survey, 2007). The suspended solids constituent is commonly analyzed using a subsample, whereas suspended sediment concentrations are commonly analyzed using a whole sample aliquot.

T

Total concentration The concentration of a constituent regardless of its form (dissolved or suspended) in a sample (U.S. Geological Survey, 2007).

Total maximum daily load (TMDL) The sum of the individual wasteload allocations and load allocations for point sources, nonpoint sources, and natural background sources. A margin of safety is included with the two types of allocations so that any additional loading, regardless of source, would not produce a violation of water-quality standards (U.S. Environmental Protection Agency, 1991). TMDLs can be expressed in terms of mass per time, toxicity, or other appropriate measures that relate to a state's water-quality standard (U.S. Environmental Protection Agency, 2007).

Toxic Relating to harmful effects to biota including death, disease, behavioral abnormalities, cancer, genetic mutations, physiological malfunctions (including malfunctions in reproduction) or physical deformations, in such organisms (or their offspring) caused by a substance or contaminant (U.S. Environmental Protection Agency, 1994; National Water-Quality Monitoring Council, 2006).

Trace element An element found in only minor amounts (concentrations less than 1.0 milligram per liter) in water or sediment; includes arsenic, cadmium, chromium, copper, lead, mercury, nickel, and zinc (U.S. Geological Survey, 2007).

Transformation A function that serves to modify a response or factor, usually motivated to make a particular model fit better or be more easily interpreted. The most common transformation is to replace a variable by its logarithm (National Institute of Standards and Technology, 2003).

U

Uncertainty A term for the fuzzy concept of qualifying statement of what is known or concluded with quantitative statements of probability. Uncertainty usually has two aspects: that created by the experimental (for example, observational) error associated with taking observations, and that implied by the use of imperfect models (National Institute of Standards and Technology, 2003). Uncertainty occurs because of a lack of knowledge. It is not the same as variability. For example, a risk assessor may be very certain that different people drink different amounts of water but may be uncertain about how much variability there is in water intakes within the population. Uncertainty can often be reduced by collecting more and better data, whereas variability is an inherent property of the population being evaluated. Variability can be better characterized with more data but it cannot be reduced or eliminated. Efforts to clearly distinguish between variability and uncertainty are important for both risk assessment and risk characterization (U.S. Environmental Protection Agency, 2006). Uncertainty includes parameter uncertainty (measurement errors, sampling errors, systematic errors), model uncertainty (uncertainty due to necessary simplification of real-world processes, misspecification of the model structure, model misuse, use of inappropriate surrogate variables), and scenario uncertainty (descriptive errors, aggregation errors, errors in professional judgment, incomplete analysis) (U.S. Environmental Protection Agency, 1997).

Urbanized area An urbanized area is a land area comprising one or more places —central place(s)—and the adjacent densely settled surrounding area—urban fringe—that together have a residential population of at least 50,000 and an overall population density of at least 1,000 people per square mile. The Bureau of the Census' general definition of an urbanized area, based on population and population density (Patricia Cazenas, Federal Highway Administration, Office of Natural and Human Environment, written comun., 2007).

V

Variability Variability refers to observed differences attributable to true heterogeneity or diversity in a population or exposure parameter. Variability is usually not reducible by further measurement or study (but can be better characterized) (U.S. Environmental Protection Agency, 1997,). The extent to which results from multiple results of the same measurement yield differing results. Variability may be inherent in a measuring instrument or in the sampled material (Jones, 2003).

Volatile organic compounds (VOCs) Organic chemicals that have a high vapor pressure relative to their water solubility. VOCs include components of gasoline, fuel oils, and lubricants, as well as organic solvents, fumigants, some inert ingredients in pesticides, and some by-products of chlorine disinfection (U.S. Geological Survey, 2007).

W

Waters of the United States All natural waters in the United States that may conceivably affect interstate or international commerce and the tributaries and components of these waters (even if intermittent or ephemeral) as defined in 40 CFR 232.2 (U.S. Environmental Protection Agency, 1994).

Water-quality criteria Numeric and narrative criteria defining water-quality objectives. Numeric criteria are scientifically derived ambient concentrations developed by the USEPA or states for various pollutants of concern to protect human health and aquatic life. Narrative criteria are statements that describe the desired water-quality goal (U.S. Environmental Protection Agency, 1991).

Water-quality data Chemical, biological, and physical measurements or observations of the characteristics of surface and ground waters, atmospheric deposition, potable water, treated effluents, and wastewater and of the immediate environment in which the water exists (National Water-Quality Monitoring Council, 2006).

Water-quality information Derived through analysis, interpretation, and presentation of water-quality and ancillary data (National Water-Quality Monitoring Council, 2006).

Water-quality monitoring An integrated activity for evaluating the physical, chemical, and biological character of water in relation to human health, ecological conditions, and designated water uses (National Water-Quality Monitoring Council, 2006).

Water-quality standard A law or regulation that consists of the beneficial designated use or uses of a water body, the numeric and narrative water-quality criteria that are necessary to protect the use or uses of that particular water body, and an antidegradation statement (U.S. Environmental Protection Agency, 1991).

Water-quality-transport curve Regression relations used to estimate constituent concentrations from measured or estimated water discharge.

Watershed A drainage area or basin in which all land and water areas drain or flow toward a central collector such as a stream, river, or lake at a lower elevation (U.S. Geological Survey, 2007).

Y

Yield The mass of material or constituent transported by a river in a specified period of time divided by the drainage area of the river basin (U.S. Geological Survey, 2007).

Granato and others—**Methods for Development of Planning-Level Estimates of Water Quality at Unmonitored Stream Sites in the Conterminous United States**—FHWA-HEP-09-003