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-09-004
2009

Highway-Runoff Database (HRDB Version 1.0): A Data Warehouse and Preprocessor for the Stochastic Empirical Loading and Dilution Model

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.

The FHWA, the Transportation Research Board and the National Cooperative Highway Research Program have repeatedly identified a national highway-runoff database as a primary environmental research need over the past decade. The Highway Runoff Database and its graphical user interface provide a tool for defining the quantity and quality of highway runoff at monitored sites and estimating runoff characteristics at unmonitored sites. This information is vital for assessing the potential for adverse effects of runoff on receiving waters throughout the Nation. Use of this database as a data warehouse should improve the usefulness and availability of runoff-monitoring results for all transportation agencies. Ready availability of this highway-runoff data in a standard format and the ease of use of the graphical user interface should provide information to improve project delivery without compromising environmental 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

Quality Assurance Statement

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipients Catalog No.
FHWA-HEP-09-004		

4. Title and Subtitle
Highway-Runoff Database (HRDB Version 1.0): A Data Warehouse and Preprocessor for the Stochastic Empirical Loading and Dilution Model

5. Report Date
June 2009

6. Performing Organization Code

7. Authors
Gregory E. Granato and Patricia A. Cazenas

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 Street SE
Washington, DC 20590

13. Type of Report and Period Covered
Final Report
October 2003–September 30, 2009

14. Sponsoring Agency Code

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

16. Abstract

The highway-runoff database (HRBD) was developed by the U.S. Geological Survey, in cooperation with the Federal Highway Administration, to serve as a data warehouse for current and future highway-runoff data sets. The database can be used to document information about a data set, monitoring site(s), highway-runoff data (including precipitation, runoff, and event-mean concentrations of water-quality constituents), quality-assurance and quality-control data, and sediment-quality data. The HRDB provides information and data that may be used to assess potential effects of highway runoff on receiving waters and the need for management measures to mitigate the potential for adverse effects on receiving waters. The HRDB application also was developed to serve as a data preprocessor for the Stochastic Empirical Loading and Dilution Model (SELDM). The HRDB application, which is the graphical user interface and associated computer code, can be used to facilitate estimation of statistical properties of runoff coefficients, runoff-quality statistics, and relations between water-quality variables in highway runoff from the available data. This report is a manual for step-by-step use of the HRDB graphical user interface and it documents the HRDB design and database application.

Many highway-runoff studies have been done over the years to collect necessary data, but the data have not been available in a consistent and accessible electronic format. The HRBD currently includes 37 tables with data for 39,713 event mean concentration (EMC) measurements (including over 100 water-quality constituents) from 2,650 storm events, monitored at 103 highway-runoff monitoring sites in the conterminous United States, as documented in 7 selected highway-runoff data sets. These data include the 1990 FHWA data compilation and results from 6 other data sets collected during the period 1993–2005.

17. Key Words	18. Distribution Statement
Highway runoff, runoff coefficient, event mean concentration, water quality, sediment quality, Stochastic Empirical Loading and Dilution Model	No restrictions.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	57	

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)**				
^oF	Fahrenheit	5 (F-32)/9 or (F-32)/1.8	Celsius	^oC	^oC	Celsius	1.8C+32	Fahrenheit	^oF
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

Table

Plate

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

ADT	average daily traffic
AML	adjusted maximum likelihood
ASCE	American Society of Civil Engineers
BMP(s)	Best Management Practice(s)
CALTRANS	California Department of Transportation
CD-ROM	computer disk, read only memory
EMC	event mean concentration
E/R	entity/relation
FHWA	Federal Highway Administration
FK	foreign key
GPS	global positioning system
IE	Information Engineering
KTRLine	Kendall-Theil robust line
MDL	Multiple Detection Limit
NCHRP	National Cooperative Highway Research Program
NDAMS	National Highway Runoff Data and Methodology Synthesis
NPDES	National Pollution Discharge Elimination System
NURP	Nationwide Urban Runoff Program
ODBC	open-database-connectivity
OLE	object linking and embedding
PCODE	U.S. Environmental Protection Agency parameter code
PK	primary key
QA/QC	Quality assurance and quality control
ROS	regression on order statistics
R_v	runoff coefficient (ratio of runoff to precipitation volumes)
SELDM	Stochastic Empirical Loading and Dilution Model
SSC	suspended-sediment concentrations
TARP	Technology Acceptance and Reciprocity Partnership
TMDL(s)	Total Maximum Daily Load(s)
TSS	total suspended solids
TRB	Transportation Research Board
URL	uniform resource locator (Internet or Web address)
USEPA	U.S. Environmental Protection Agency
WERF	Water Environment Research Foundation
VBA	Visual Basic for Applications

Highway-Runoff Database (HRDB Version 1.0): A Data Warehouse and Preprocessor for the Stochastic Empirical Loading and Dilution Model

By Gregory E. Granato and Patricia A. Cazenas

Abstract

The highway-runoff database (HRDB) was developed by the U.S. Geological Survey, in cooperation with the Federal Highway Administration (FHWA), to serve as a data warehouse for current and future highway-runoff data sets. The database can be used by transportation agencies and researchers as a data warehouse to document information about a data set, monitoring site(s), highway-runoff data (including precipitation, runoff, and event mean concentrations of water-quality constituents), quality-assurance and quality-control data, and sediment-quality data. Information and data about the quantity and quality of highway runoff can be used to document runoff properties (flows, concentrations, and loads) at monitored sites and to estimate these runoff properties for unmonitored sites with similar characteristics. The HRDB provides information and data that may be used to assess potential effects of highway runoff on receiving waters and the need for management measures to mitigate the potential for such adverse effects.

Many highway-runoff studies have been done over the years to collect necessary data, but the data have not been available in a consistent and accessible electronic format. The HRDB currently includes 37 tables with data for 39,713 event mean concentration (EMC) measurements (including over 100 water-quality constituents) from 2,650 storm events, monitored at 103 highway-runoff monitoring sites in the conterminous United States, as documented in 7 selected highway-runoff data sets. These data include the 1990 FHWA runoff-quality model data compilation and results from 6 other data sets collected during the period 1993–2005.

The HRDB application also was developed to serve as a data preprocessor for the Stochastic Empirical Loading and Dilution Model (SELDM). SELDM is a water-quality model that is designed to help estimate runoff flows, concentrations, and loads from highways and in receiving waters at unmonitored sites based on site characteristics. The HRDB application, which is the graphical-user interface and associated computer code, can be used to facilitate estimation of statistical properties of runoff coefficients, runoff-quality statistics, and relations between water-quality variables in highway runoff from the available data. The database application facilitates retrieval and processing of the available highway-runoff data.

This report is a manual for step-by-step use of the HRDB graphical-user interface and it documents the HRDB design and database application. The highway-runoff data in the database is discussed to provide an overview of the database contents and examples of the potential use of such data. Some basic information about database design and implementation in Microsoft Access is provided. The data structures and table definitions that constitute the database contents are described in this report, on a database design diagram, and in a data dictionary on the accompanying CD-ROM. The program code, written in Microsoft Visual Basic for applications, is documented in this Microsoft Access database file on the accompanying CD-ROM. The report also documents operational issues and procedures for current and future use of this database and the database application.

Introduction

Knowledge of the properties of highway runoff, including event mean concentrations (EMC) of water-quality constituents, runoff flows, and runoff loads, is important for decision makers, planners, and highway engineers to assess and mitigate possible adverse effects of highway runoff on the Nation's receiving waters (Bank, 1993; Transportation Research Board 2002; Granato, Zenone, and Cazenas, 2003). Data and information about precipitation and the quality and quantity of highway runoff from sites with different highway-design characteristics, traffic volumes, and surrounding land uses help define variations in runoff quality from site to site. Data and information from different areas of the country may be used to characterize the quality of highway runoff as a function of regional variations in fuel formulations, emission standards, construction and maintenance practices, and soil geochemistry. Highway-runoff data also are necessary to assess the need for and potential effectiveness of management measures, (such as structural best management practices (BMPs), to mitigate the potential for any adverse effects of runoff on receiving waters. Finally, such data are necessary to formulate planning-level estimates of runoff quality for existing or planned sites for which monitoring data are unavailable. Organization and centralization of highway-runoff data from various sources has consistently been identified as a high-priority environmental-research need by the Federal Highway Administration (FHWA), the Transportation Research Board (TRB), and the National Cooperative Highway Research Program (NCHRP) (Bank, 1993; Transportation Research Board 1993; 1996a; 1996b; 1997; 2002; Venner and others, 2004).

Publication of the 1990 FHWA runoff-quality model with data from a number of data-collection studies was the culmination of the FHWA runoff-quality research conducted during the 1970s and 1980s (Driscoll and others, 1990 a,b,c,d). The 1990 FHWA runoff-quality model was based on this older, available runoff-quality data and the assumption that concentrations of water-quality constituents in receiving waters were equal to zero. By the mid-1990s, however, it was recognized that the existing data and modeling methods would reach obsolescence as time went on because of changes that have occurred since the original field monitoring studies were completed (Bank and others, 1996). Changes in highway construction and maintenance activities (such as the use of pulverized rubber tires in pavement mixtures), and automobile technology (such as the disappearance of leaded fuel, continuing improvements in catalytic converters, and a technological trend from asbestos to organo-metallic brake pads) may affect the quality of highway runoff. Changes in atmospheric deposition and other ambient sources of pollution from surrounding land uses also could affect the quality of highway runoff. These and other changes may substantially alter the quality of runoff and the potential effects of this runoff on some receiving waters. In addition, as a result of the Total Maximum Daily Loads (TMDLs) process, regulators and decision makers have become increasingly aware of the importance of considering the quality of upstream receiving waters for examining potential effects of runoff from highways and other land uses.

Regional and National Highway-Runoff Information Needs

Recognition of need for available, consistent, and technically sound runoff-monitoring data has led to several standardization efforts by federal and state agencies, universities, and highway practitioners. This need was highlighted by the findings of the FHWA National Highway Runoff Data and Methodology Synthesis (NDAMS) (Granato and others, 1998; Granato, 2003). Results of the NDAMS study indicate that knowledge of the details of highway-runoff studies is not persistent or pervasive and that detailed data and documentation for studies more than 5 years old often are unobtainable because of changes in personnel and computer systems (Granato, Dionne, Tana, and King, 2003). The NDAMS study cataloged and reviewed a sample of 250 highway-runoff studies and indicated that few highway-runoff monitoring reports available at that time would meet current documentation standards and data-quality requirements (Granato, 2003). In response to these information needs, the NDAMS project produced a compilation of chapters, each written by subject-matter experts, to define requirements for defensible data sets for each facet of a highway-runoff monitoring study (Granato, Zenone, and Cazenas, 2003). The FHWA also published a guidance manual for monitoring highway-runoff quality to help standardize methods and results of highway-runoff monitoring studies (Strecker and others, 2001).

Similarly, other organizations have documented an increased emphasis on data standardization, documentation, quality, defensibility, and availability. On a national scale, the U.S. Environmental Protection Agency (USEPA), the American Society of Civil Engineers (ASCE), and the Water Environment Research Foundation (WERF) published a guidance manual for BMP performance monitoring in an effort to compile the data necessary to improve BMP selection and design for inclusion in the International BMP database (Strecker and others, 2002). On a regional scale, the Technology Acceptance and Reciprocity Partnership (TARP), which includes environmental monitoring and regulatory agencies from California, Massachusetts, New Jersey, Pennsylvania, and Virginia, also has established protocols for monitoring runoff, documenting methods and data, and interpreting the results of studies of BMPs (Technology Acceptance and Reciprocity Partnership, 2001). At the state level, the California Department of Transportation (CALTRANS) developed and published a set of stormwater monitoring protocols to collect and store data of known accuracy and precision (California Department of Transportation, 2000). The CALTRANS manual was written so that data would be suitable to support the CALTRANS stormwater management program, to comply with various regulatory and legal requirements, and to be scientifically defensible in a range of other potential applications (California Department of Transportation, 2000). Information needs identified by CALTRANS include characterization of the quality and quantity of discharges, evaluation of BMP performance, runoff modeling, comparisons to other studies, and assessments of highway-contributions to receiving water loadings. These data-collection programs are beneficial but none are focused on national highway-runoff information needs.

A recent study by the NCHRP (Venner and others, 2004) concluded that a national highway-runoff database, available in the public domain, was necessary to document the results of monitoring efforts to characterize the quality of runoff from operating highways. This NCHRP study concluded that a database, which included a structure to record detailed results of runoff-monitoring studies (such as is found in the International BMP Database or the CALTRANS proprietary database) as well as the bibliographic and data-quality information in the NDAMS database was necessary to further highway-runoff research. The International BMP Database does not have a bibliographic component that identifies source documents for the data. Identification of source documents facilitates investigation of the study design, the field and laboratory methods used, and the availability of quality-assurance and quality-control data. Examination of the reports that document detailed methods and results of water-quality studies commonly reveal the specific site characteristics, individual methods descriptions, and the results of quality-assurance and quality-control measures that are necessary to properly use such data. Furthermore, the International BMP Database accepts only highway and urban-runoff characterization data collected as part of comparative (input versus output) BMP studies. The design of the CALTRANS proprietary database is well suited for documenting CALTRANS monitoring efforts, but that database contains many types of data and is complex. Both the CALTRANS proprietary database and the International BMP Database are complex enough to be supported and maintained by professional database administrators. Therefore, a relatively simple data structure was needed to store available highway-runoff data, to provide researchers with a common data format to record results from current and future runoff studies, and to facilitate the export of data and summary statistics for further analysis of runoff properties.

To address evolving information needs the U.S. Geological Survey (USGS), in cooperation with the FHWA, began to develop a new water-quality model, known as the Stochastic Empirical Loading and Dilution Model (SELDM), to supersede the 1990 FHWA runoff-quality model. Runoff coefficients and EMC statistics are used with SELDM to generate random populations of runoff volumes, concentrations, and loads from regional precipitation statistics and site characteristics by use of Monte Carlo methods. This information may be used to estimate runoff quantity and quality based on site characteristics, and to predict potential effects of highway runoff on receiving waters. Proper application of such a model, however, requires technically sound statistical estimates of the quality and quantity of runoff and receiving waters upstream of the highway outfall. Such statistical estimates require technically sound and well-documented data and statistically valid estimation methods appropriate for the data. As SELDM was developed, it was realized that use of the model, as well as other analyses and applications of highway-runoff data, would be greatly facilitated by a database for complete and comprehensive storage, retrieval, and analysis of these data in a consistent format. Thus, a data warehouse was created to document data and information from available highway-runoff monitoring studies.

Purpose and Scope

This report is a manual for the HRDB application and describes the use, design, and contents of the application. The HRDB application is designed as a data warehouse to document data and information from available highway-runoff monitoring studies and as a preprocessor for highway-runoff data for use in the SELDM application. The availability of highway-runoff data provides the basis for defining runoff quality and quantity at monitored sites and predicting runoff quality and quantity at unmonitored sites. The data that were used to develop the 1990 FHWA runoff-quality model (Driscoll and others 1990a, b, c, d) are included as a basis for comparison with newer data. Additional data from six newer highway-runoff data sets that were available with a substantial amount of supporting documentation are included as an initial update to the earlier data set.

The HRDB application also is designed to be a preprocessor for use with SELDM. Most common data-manipulation tasks can be accomplished with the graphical-user interface of the HRDB or by use of several predefined queries with only a cursory knowledge of Microsoft Access. The database application provides standard and robust estimates of population statistics for highway-runoff data. The procedures for manipulating data in the database application are described, and step-by-step use of the application's graphical-user interface is illustrated.

Information about the design and implementation of the application and underlying database are provided to facilitate future use and modification of the highway-runoff database application. The program code, written in Microsoft Visual Basic for applications, is documented in the Microsoft Access database file on the CD-ROM accompanying this report. Some basic information about database design and implementation in Microsoft Access is provided. The implementation and design portions of this report, however, are written with the assumption that potential users who would be making design changes would have a working knowledge of Microsoft Access and some background in the design or use of relational databases. Information and training on the use of Microsoft Access is widely available and can be located on the Internet. Information about data models and relational database-design concepts are available in many books (for example, Fleming and von Halle, 1989; Hernandez, 1997; Roman, 1997), and in the Federal data-modeling standard document FIPS 184 (National Institute of Standards and Technology, 1993).

The primary function of this document is intended to be as a manual for the HRDB. Although the presented order of topics is from subject data, to use of the graphical-user interface application, to design and implementation of the underlying database, some readers may prefer to read the document in a different order. If the reader needs highway-runoff data or statistics, the first two sections after the introduction should provide the necessary information. If the reader needs data not provided by the standard choices in the database application, then it will be necessary to understand the database design and contents. If the reader needs to add data, extend the database, or act as an administrator for an updated version of the database, then information about operational issues and procedures also is necessary.

Highway-Runoff Data

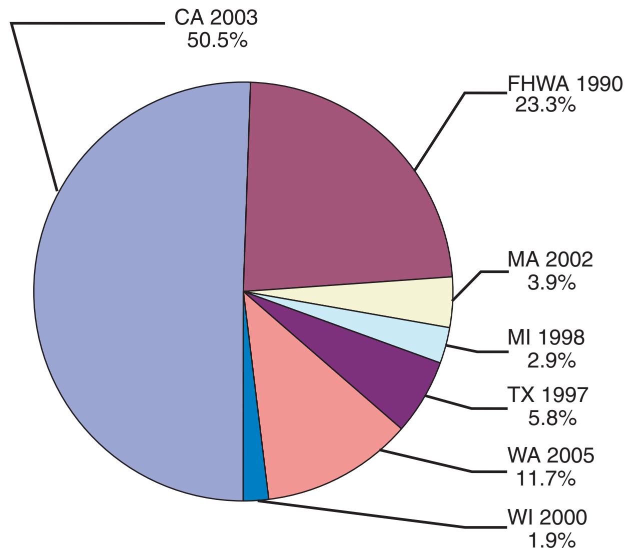
Information and data about the quantity and quality of highway runoff are necessary to assess the potential effect of highway runoff on receiving waters and the need for management measures to mitigate the potential for these effects. Selected data sets from previous studies form the core of a future FHWA highway-runoff data warehouse, provide an initial data set for use with SELDM, and provide data used to develop and test the database application and the underlying data model. Information about the data included in this version of the database is summarized, and selected properties of highway-runoff data are explored. Driscoll and others (1990c) documented a detailed analysis of properties of highway runoff, factors that influence highway runoff quantity and quality, and approaches to predictive modeling. This summary provides an overview of an updated data set that may be used for such an analysis.

Currently, the database includes data from 7 highway-runoff data sets with 103 sites, 2,650 storms, and 39,713 individual stormwater-quality measurements (fig. 1). Data from the 1990 FHWA runoff-quality model “working database,” which represents a compilation of previous studies (Driscoll and others, 1990c; d) are included to supplement and to provide a basis for comparison with newer data sets. The California data set currently is the largest highway-runoff data set collected, processed, analyzed, and recorded in a robust and consistent data-quality system (California Department of Transportation, Office of Environmental Engineering, 2000; 2002; 2003a; b; c; d; 2004). Highway-runoff data from Massachusetts represents results from a BMP characterization study (Smith, 2002). The Wisconsin study (Waschbusch, 2003) documents highway-runoff quality with and without street sweeping. The Washington State data sets include highway runoff characterization data for National Pollution Discharge Elimination System (NPDES) permits (Tetra Tech, Inc., 2002; 2003; 2004; Washington State Department of Transportation, Environmental Services Office, 2001, 2002, 2003, 2004) and BMP monitoring data (Taylor Associates, Inc., 2002a; 2002b). The Michigan Department of Transportation data are results from a highway stormwater-runoff characterization study (CH2MHill Inc., 1998). The Texas data set represents results from a characterization study (Barrett and others, 1995, 1996) and a BMP study (Walsh and others, 1997). The storm events in the highway-runoff database span a period of three decades from 1975 to 2005 (fig. 2). Although there are 103 data-collection sites, 24 have data collected before 1986, 52 are distributed in California, and the remaining 27 are clustered in 5 states (fig. 3).

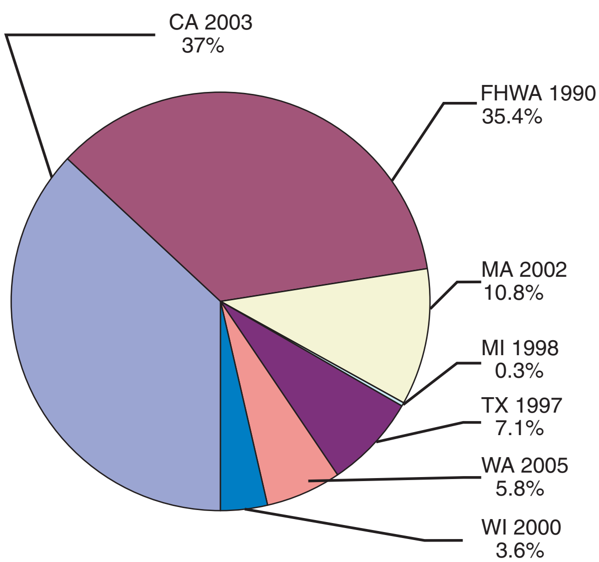
A. Count of sites, storm events, and event-mean concentration (EMC) values

Highway-Runoff Data Set	Count of Sites	Count of Storms	Count of EMCs
Federal:			
FHWA 1990	24	937	8,039
State:			
CA 2003	52	981	26,104
MA 2002	4	285	1,236
MI 1998	3	9	198
TX 1997	6	187	1,925
WA 2005	12	155	1,486
WI 2000	2	96	725
Sum	103	2,650	39,713

B. Percentage (%) of sites in each data set



C. Percentage (%) of storm events in each data set



D. Percentage (%) of event mean concentration values in each data set

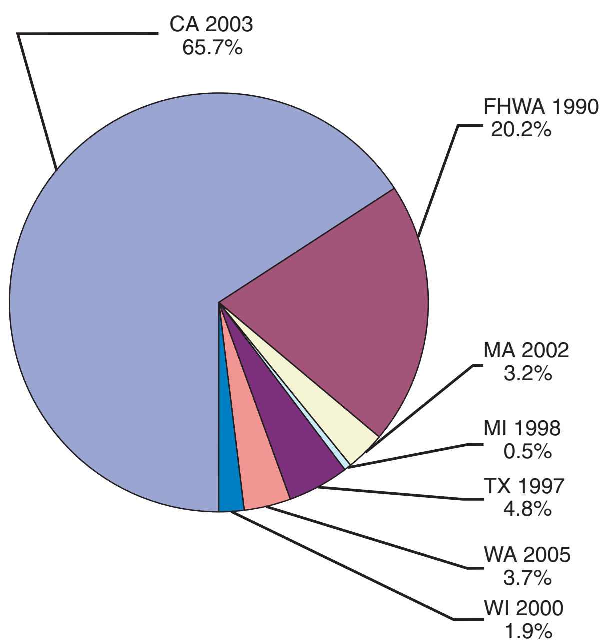


Figure 1. Summary of the highway-runoff data including *(A)*, a count of sites, storms, and event mean concentration values in the database, and the percentage of; *(B)*, sites; *(C)*, storm events; and *(D)*, event mean concentration (EMC) values in each highway-runoff data set.

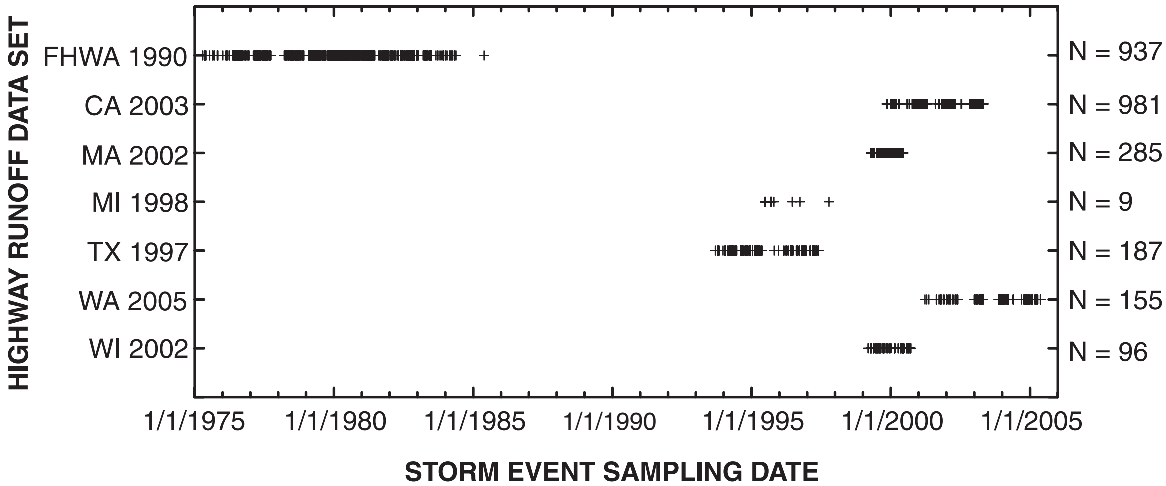


Figure 2. The temporal distribution of storm-event sampling dates for each data set in the highway-runoff database. (N, number of storm events in each data set).

Figure 3. Index map showing highway-runoff monitoring stations from the working database of the 1990 Federal Highway Administration compilation and new sites from six highway-runoff data sets in the conterminous United States (geographic projection).

Highway Runoff Coefficients

Runoff coefficients commonly are used to relate the amount of precipitation that may occur in a given storm to the average amount of runoff generated from a highway site during that storm. In the nationwide urban runoff program (NURP), Athayde and others (1983) defined the runoff coefficient (R_v) as the ratio of runoff volume to rainfall volume, and determined that the variation in R_v at individual study sites was a random variable that was well-defined by a lognormal distribution. Driscoll and others (1990c) also concluded that runoff coefficients from individual sites could be characterized as random, lognormal variables. Runoff coefficients are theoretically bounded between zero (no runoff) and one (100 percent of precipitation runs off). Runoff coefficients are expected to vary from storm to storm with antecedent conditions and to vary from site to site as a function of impervious area (Athayde and others, 1983; Schueler, 1987; Driscoll and others, 1990c). In practice, however, uncertainties in measurement of rainfall, runoff volumes, impervious areas, and the total contributing area for each storm can yield runoff coefficients that are greater than one (Church and others, 2003).

Runoff coefficients can be used to predict runoff volumes and runoff-constituent loads. Highway-runoff data sets commonly include a relatively small number of highway sites and a relatively few number of storms per site (Driscoll and others, 1990c). Rainfall data or estimates of rainfall statistics, however, are available throughout the nation and this information may be used to predict runoff at unmonitored sites (Driscoll and others, 1989). Researchers commonly use a regression equation to predict runoff coefficients from estimates of the fraction (or percentage) of impervious area based on the average runoff coefficient from each site (Athayde and others, 1983; Schueler, 1987; Driscoll and others, 1990c).

Average runoff coefficients commonly are used to predict runoff volumes because of the uncertainties in individual measurements (Strecker and others, 2001; Church and others, 2003). Of the 103 sites in the highway-runoff database, 84 sites have the rainfall measurements, runoff measurements, and drainage-area estimates that are necessary to calculate runoff coefficients for a given site; 83 sites have an estimated impervious area (fig. 4). High variability in runoff coefficients at a given site is expected from storm to storm. Variability in antecedent conditions, rapid changes in precipitation intensity and runoff, and uncertainty in measurement methods can account for high variability in runoff coefficients. For example, 39 sites have individual runoff coefficients that vary by more than an order of magnitude, and 53 sites have maximum runoff coefficients that are substantially greater than one.

Of the 84 sites with the information necessary to calculate runoff coefficients, 24 sites have an average runoff coefficient that is greater than one, and 9 sites have an average runoff coefficient that is abnormally low (less than 50 percent of what would be expected based on impervious area). Systematic bias in the entire population of values indicates a problem in the drainage-area estimate. Precise estimates of drainage area are difficult in small highway catchments that are the subject of water-quality investigations (Strecker and others, 2001). It is difficult to accurately delineate a small low-slope catchment, because small surface features have an inordinate effect on drainage patterns in these catchments. Vehicles can track water along the roadway and spray water off the pavement and into the air. For example, bias in the runoff coefficients at the sites in Massachusetts are caused by periodic bypass flows from neighboring drainage areas along ruts in the roadway and along the road edge around neighboring catch basins to these sites, which are at a low spot in the road. These bypass flows occur during periods of high-intensity rainfall during some storm events and increase the effective drainage area of the monitored subcatchment at these sites (K.P. Smith, U.S. Geological Survey, oral commun., 2005). Therefore, estimates of runoff coefficients must be adjusted so that the maximum runoff coefficient does not exceed one to eliminate potential mass-balance errors in runoff estimates made from precipitation records for an entire catchment.

The 1990 FHWA runoff-quality model study used data from 18 sites (from a total of 789 storms) to determine that (1) the runoff coefficient commonly is independent of the total rainfall volume for a given storm, (2) runoff coefficients for different storm events at a given site vary lognormally, and (3) among different sites, the impervious fraction of the contributing drainage area is a satisfactory explanatory variable for the expected runoff coefficient (Driscoll and others, 1990c). Examination of site characteristics for the 83 sites with rainfall data, runoff data, drainage area, and the impervious fraction indicates that many of the sites with lower impervious fractions tend to have higher drainage areas (fig. 4a). One may expect reduced variability in storm-to-storm runoff coefficients at each site with increasing impervious fraction, because paved areas commonly are designed to convey rather than retain precipitation. The relatively constant coefficient of variation (COV) values over the range of impervious fractions (fig. 4b) in the data set, however, probably are an artifact of the distribution of drainage areas among the different sites. The larger drainage areas of the sites with lower impervious fractions potentially reduce storm-to storm variations in measured values. The COV of the smaller sites with higher impervious fractions potentially reflect the effect of variable contributing areas from storm to storm.

The regression analysis from the 1990 FHWA runoff-quality model study, was based on the average runoff coefficient from the largest 15 sites with various impervious fractions. This regression analysis indicated that the equation for the average runoff coefficient has a slope of about 0.7 (times the impervious fraction) and an intercept of about 0.1. In the current study, regression analysis of the average runoff coefficients for 44 sites in the highway-runoff database (including sites from the 1990 FHWA runoff-quality model study) that have reasonable average runoff coefficients indicates a slope of about 0.67 (times the impervious fraction)

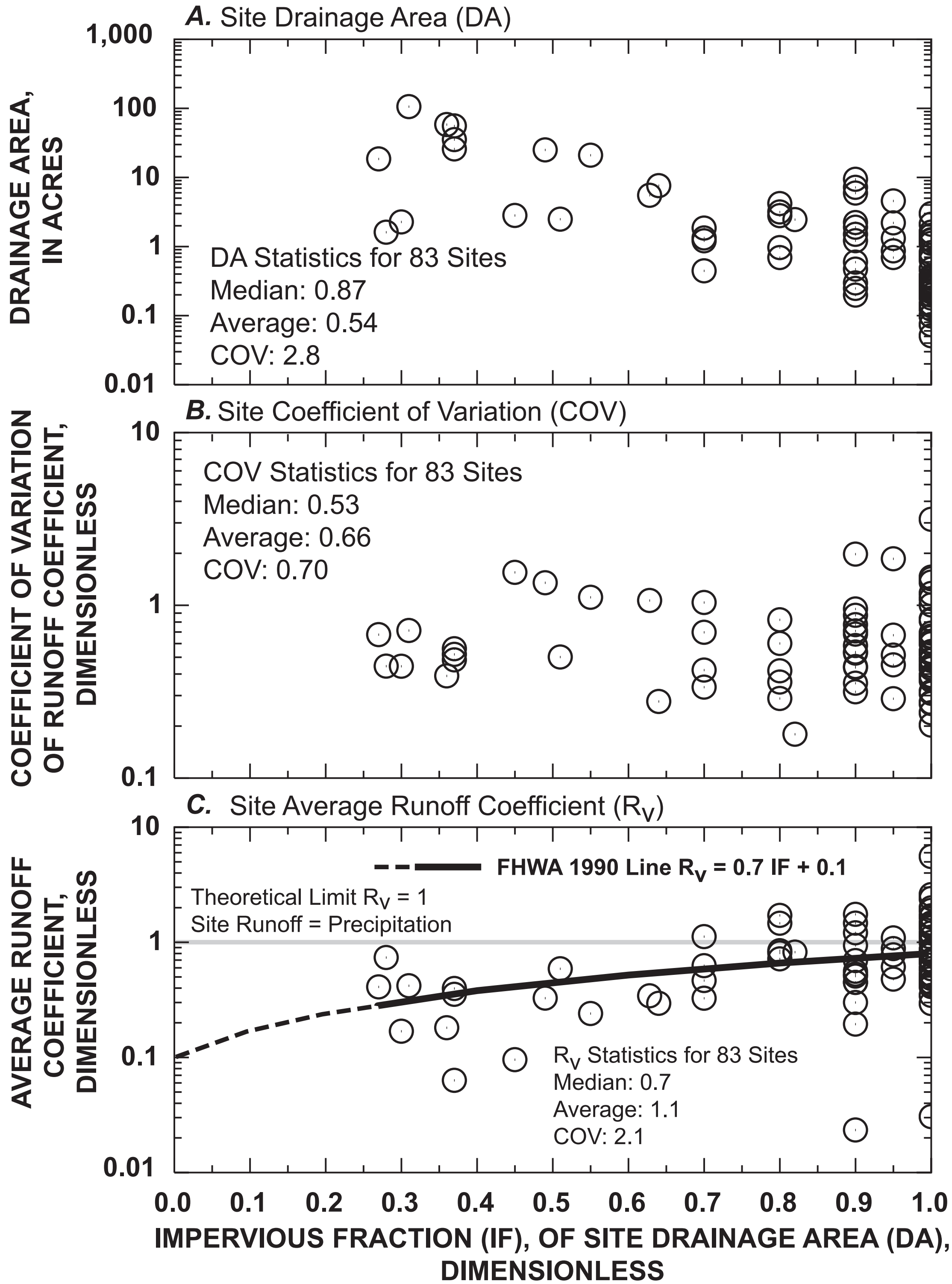


Figure 4. Information and statistics including *(A)*, drainage area estimates; *(B)*, the site coefficient of variation of runoff coefficient values from individual storms; and *(C)*, the site average of runoff coefficient values from individual storms at each of the 83 highway-runoff monitoring sites that have precipitation, runoff, impervious fraction, and drainage area data.

and an intercept of about 0.08. The slope and intercept from the original FHWA equation for average runoff coefficients is well within the 95-percent confidence interval of the new equation and is well within the (considerable) scatter of the site-average runoff coefficients around the regression line. Thus, continued use of the 1990 equation for planning-level estimates of runoff volumes is supported by the current analysis with more highway sites (fig. 4c).

Event Mean Concentration Data

The EMC is operationally defined as the total water-quality-constituent mass discharged during a storm divided by the total volume of the runoff and is, therefore, the average pollutant concentration present in the total volume of runoff from a storm event (Athayde and others, 1983; Schueler, 1987; Driscoll and others, 1990c). EMCs can be derived by mathematical computation of discrete measurements of concentration and runoff, or by analysis of a single flow-weighted composite sample collected during a storm (Athayde and others, 1983; Schueler, 1987; Driscoll and others, 1990c; U.S. Environmental Protection Agency, 1992; Strecker and others, 2001; 2002). Because analytical costs for discrete instorm samples effectively reduce the number of storms that can be sampled, many studies produce data based on flow-weighted composite samples unless the research is focused on instorm processes (Driscoll and others, 1990c; Strecker and others, 2002).

The highway-runoff database includes 39,713 EMC measurements from 2,650 storm events, monitored at 103 highway-runoff monitoring sites in the conterminous United States, as documented in 7 highway-runoff quality data sets (fig. 1). These EMC measurements include measurements for 116 different water-quality constituents and water-quality properties (such as oxygen demand, solids, specific conductance, temperature, and pH). These water-quality measurements include 17,810 trace-metal EMCs; 9,267 physical property EMCs; 6,002 nutrient EMCs; 3,375 major inorganic constituent EMCs; 2,987 organic constituent EMCs; and 272 other EMC measurements. Several of the data sets have associated quality-assurance and quality-control (QA/QC) data that are not entered in the database and are not included in these totals. Examination and entry of the QA/QC data was beyond the scope of the current study because these data would require additional scrutiny and must be entered in a separate table in the database.

Robust estimates of population statistics for highway-runoff volumes and EMCs are necessary to develop planning-level estimates of the concentrations and loads of these properties and constituents in runoff at unmonitored sites throughout the Nation. Data for concentrations and loads of highway runoff indicate the expected quality of runoff at a given site and define the potential for adverse effects caused by discharge of highway runoff in a watershed. The need for management measures to mitigate the potential for adverse effects of runoff is determined by the probability that the runoff will have an adverse effect on receiving waters. In the 1990 FHWA runoff-quality model study, Driscoll and others (1990c) segregated available highway-runoff data (8,039 EMC measurements for 19 constituents from 24 sites) into “Rural” and “Urban” sites based on traffic density with 30,000 vehicles per day as the classification criteria. They found that the sites with higher traffic density had statistically higher median concentrations and, therefore, a higher probability for water-quality exceedances. This distinction was meant to be first approximation for estimating runoff quality rather than an absolute division between sites. The original intent of the 1990 study was for the user to select summary statistics from one or more sites that best represent conditions at the site of interest (Eric Strecker, Geosyntec Consultants, oral commun., 2005). Decision makers need EMC data and statistics that can be selected on the basis of highway-site characteristics. This highway-runoff database facilitates site-by-site analysis because it includes about five times the number of monitoring sites and EMC values as the 1990 FHWA runoff-quality model compilation.

Estimates of the concentrations and loads from highway-runoff EMCs are complicated by the fact that highway-runoff quality data sets commonly include EMC measurements that are below one or more detection limits. Therefore, estimates must be made using statistical methods that are appropriate for the data. For example, Shumway and others (2002) report that 76, 43, 9, and 2 percent of measured nickel, chromium, lead, and copper concentrations, respectively, are below one or more detection limits in the California Department of Transportation highway-runoff data set. A recent summary of methods used to handle such data (Helsel, 2005) indicates that systematic and scientifically defensible methods are necessary to evaluate population statistics in a quantitative manner. Helsel (2005) also states that simple substitution methods, which have been advocated in some regulatory settings (U.S. Environmental Protection Agency, 1998; U.S. Army Corps of Engineers, 1998), may bias statistics and will vary as a function of the substitution value. In the 1990 FHWA runoff-quality model study, Driscoll and others (1990c; d) identified detection-limit issues as a potential problem but did not identify which values were censored in their working or master data sets. However, detection limits were addressed in the 1990 FHWA runoff-quality model study by use of regression on order statistics (ROS) to estimate the standard deviation of the entire population, and by use of the median to estimate the mean of log-transformed values under the assumption that all values in the data set are lognormally distributed (Driscoll and others, 1990c; d). Theoretical relations between these lognormal values and their arithmetic counterparts (Chow, 1954) were used to retransform these statistics into a mean and coefficient of variation for the data in arithmetic space.

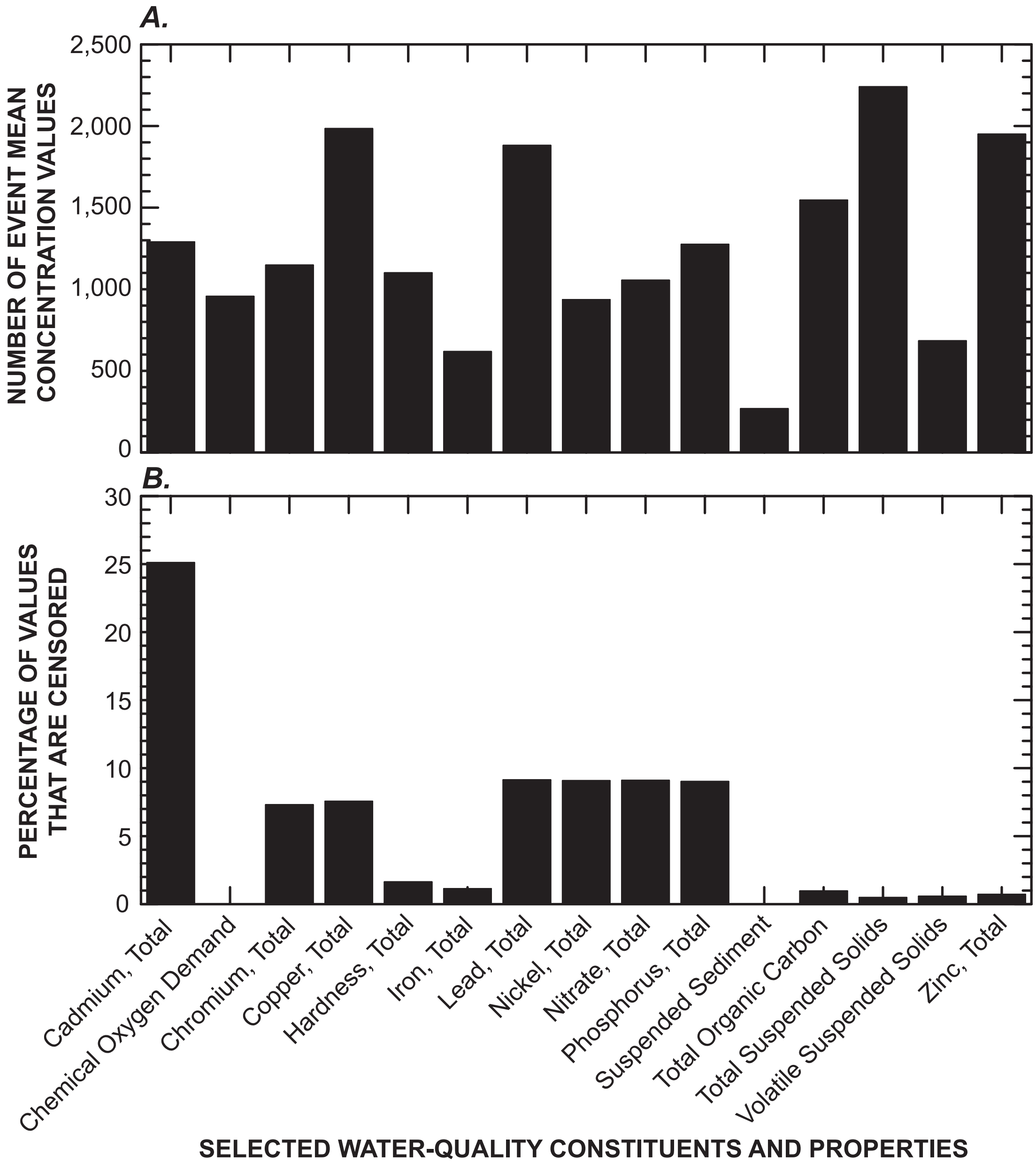


Figure 5. Number of *(A)*, event mean concentration values; and *(B)*, the percentage of these measurements that are censored values for selected water-quality constituents and properties.

About 17 percent of all the EMC measurements in the database are identified as censored values (measured or estimated below a reporting or detection limit). About 77 percent of organic constituent EMCs, 19 percent of trace-metal EMCs, and 13 percent of nutrient EMCs are identified as censored values. The number of EMC measurements and the percentage of values that are censored are shown for 15 selected water-quality constituents and properties in figure 5. More than 7 percent of EMC values are censored values for 7 of these water-quality constituents and properties, and 1 percent of EMC values are censored values for 3 of the remaining water-quality constituents and properties listed in figure 5. Therefore, methods for estimation of summary statistics for populations with censored vales (Helsel, 2005) are needed to determine planning-level estimates of highway-runoff quality.

The censored EMC measurements in the database may be from composite or discrete measurements. A censored EMC for a composite sample is a laboratory determination from analysis of an individual flow-weighted composite sample. A censored EMC from discrete measurements is the mathematical flow-weighted average of concentrations measured for two or more discrete samples from a single storm that may include one or more individual concentrations that are less than detection limits. There are no established methods for estimating the value of an EMC from discrete analyses that include one or more censored values (D.R. Helsel, U.S. Geological Survey, oral commun., 2005). In this case, the best method is to use the actual analytical reading (even though it is below detection limits) from the laboratory if such values are available. Other methods include use of surrogate parameter relations, statistical methods described by Helsel (2005), and use of the nominal detection limit for a discrete value that is used to calculate the corresponding censored EMC. Surrogate parameter relations are based on the assumption that one water-quality variable can be used to predict the concentration of another (for example, Thomson and others, 1997).

Statistical methods are theoretically rigorous but depend on the availability of enough data from within each storm to develop estimates of the mean value (D.R. Helsel, U.S. Geological Survey, oral commun., 2005). All available discrete values (from different storms) cannot be used quantitatively with statistical methods because it would be difficult to assign estimated values among the different EMCs for different storms. Use of the nominal detection limit for individual censored values among discrete measurements will provide an estimate of the censored EMC value that is conservative (biased high). If original laboratory data for subsample concentrations are not available and if the total number of uncensored EMC measurements is sufficient to use the ROS method (about 20 percent of the EMC values, Helsel, 2005) then assumptions about the concentrations of some discrete subsamples will have minimal effect on estimates of population summary statistics for all EMC values.

Technical Issues for Suspended Sediment Data

Potential problems with total suspended solids (TSS) as a measurement of sediment concentrations for monitoring highway and urban runoff, BMPs, and receiving waters have been identified (Gray and others, 2000; Smith, 2002; Bent and others, 2003; Waschbusch, 2003; U.S. Environmental Protection Agency, 2005a). Proper definition of sediment in runoff and receiving waters is critical because a review of the highway-runoff literature indicates that ecological effects in receiving waters are most likely to occur in places where runoff sediments accumulate (Buckler and Granato, 2003). The analytical methods for measuring TSS (American Public Health Association, American Water Works Association, and Water Pollution Control Federation, 1995) commonly are done with a small subsample of water that may not properly represent the full grain-size distribution of the sample (Gray and others, 2000). The method for analysis of suspended-sediment concentrations (SSC) (American Society for Testing and Materials, 2000), however, is considered more reliable because it is used to measure the dry weight of all sediment from a known volume of a water-sediment mixture (Gray and others, 2000). Gray and others (2000) indicate that because methods for TSS analysis systematically under represent the coarse fraction of the total suspended sediment in receiving waters, this method is “fundamentally unreliable for the analysis of natural-water samples.” The USGS Office of Surface Water and Office of Water Quality determined that TSS analyses are “not appropriate” for characterization of sediment concentrations (U.S. Geological Survey, 2000).

Similarly, flaws in the TSS analysis methods have been shown to under represent suspended-sediment concentrations in highway-runoff data with and without operational BMPs (street sweeping) and structural BMPs in studies that have collected paired TSS and SSC measurements (Smith, 2002; Bent and others, 2003; Waschbusch, 2003). Bent and others (2003) concluded that the systematic bias in TSS measurement also could result in substantial underestimation of the effectiveness of BMPs for removing sediment in highway runoff because the coarser sediments in the influent would not be properly characterized. The U.S. Environmental Protection Agency (2005a) also reached this conclusion from examination of potential bias in TSS measurement. The TSS method commonly is used, however, because it is a traditional method carried over from methods developed for analysis of municipal wastewater effluents. As such, TSS analysis has been specified in rules, regulations, and guidance documents for storm runoff and BMP performance. Thus, most highway and urban runoff studies include analysis of TSS measurements rather than SSC measurements to estimate the amount of sediment in runoff (Bent and others, 2003). For example, there are 2,240 TSS measurements but only 268 SSC measurements in the highway-runoff database (fig. 5). A query of the USGS National Water Information System (NWIS) Web, however, reveals that about 276,000 paired SSC and discharge measurements are available from about 7,500 surface-water-quality monitoring stations (with drainage areas less than 1,140 square miles) in the conterminous United States. In comparison, only about one-third as many measurements and monitoring sites have paired TSS and discharge measurements. Therefore, a method is needed to estimate SSC in highway runoff from available TSS data to facilitate analysis of the potential effects of sediment from runoff on receiving waters.

The 94 paired measurements of TSS and SSC available in the highway-runoff database were used to develop a surrogate-parameter relation for SSC. A log-linear regression relation was established to estimate SSC from TSS using these paired measurements (fig. 6). This relation indicates that SSC measurements are systematically higher than TSS measurements. Only about 14 percent of the paired samples have TSS concentrations that are greater than the corresponding SSC value. Similarly, Glysson and others (2000) developed regression equations from a much larger data set of paired TSS and SCC measurements (14,466 paired values) from different rivers and streams throughout the United States that indicated a systematic negative bias in TSS concentrations. Collection of SSC measurements in future

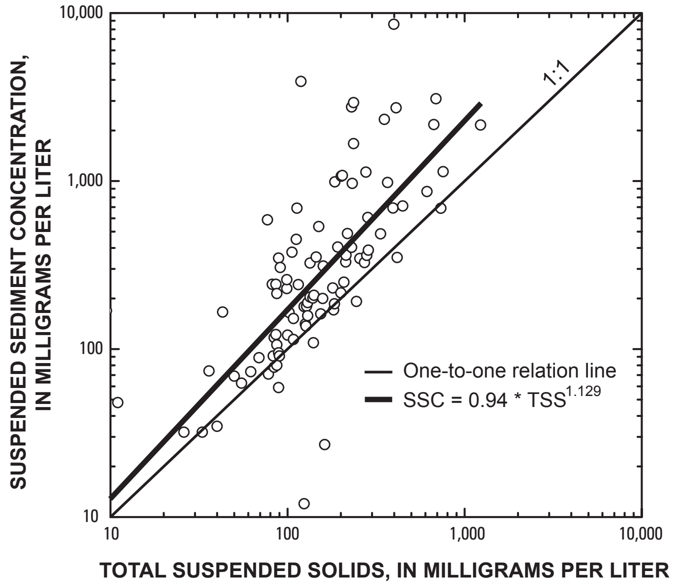


Figure 6. The relation between 94 paired total suspended solids concentrations (TSS, parameter code P00530) and suspended sediment concentrations (SSC, parameter code P80154) in comparison with a line indicating a one-to-one relation and a log-linear regression line between paired values.

highway-runoff monitoring studies may be used to provide better data for runoff analysis. Such data also may be used to refine the regression model provided herein. In the interim, the regression relation shown in figure 6 may be used to help develop planning-level estimates of concentrations and loads of SSC from highway sites that will be comparable to estimates of SSC in receiving waters.

Use of the Highway-Runoff Database Application

The HRDB application is the system of user-forms and underlying queries that constitute the graphical-user interface. This allows the user to extract data and statistics with only a minimal knowledge of the Microsoft Access software. The HRDB application was developed to facilitate use of available highway-runoff data to characterize and predict flows, concentrations, and loads of highway-runoff constituents based on site characteristics. This information and data may be used to generate planning-level estimates of runoff quality and quantity at a site of interest. Planning-level estimates of runoff quality and quantity are necessary for regulatory, planning, and design purposes (Granato, Zenone, and Cazenas, 2003). The HRDB application was designed to facilitate retrieval of the data in formats that would facilitate use of the data with other computer applications such as spreadsheets, statistical packages, the Multiple Detection Limit (MDL) Software (Helsel and Cohn, 1988; Helsel and others, 1988), and the Kendall-Theil Robust Line analysis software (Granato, 2006). The database application also is designed to facilitate calculation of the statistics necessary for analysis of highway-runoff data. The HRDB application provides the ability to export:

- water-quality data in a tab-delimited format for use with other software packages;
- water-quality data in a format for use with detection-limit software;
- paired water-quality data in a tab-delimited format for regression analysis;
- summary statistics for water-quality data with (or without) censored data; and
- information and data necessary to evaluate storm-by-storm runoff coefficients for different sites.

These five options are provided so that the user may select from all available data or a custom data set and do the analysis necessary to estimate runoff quality and flows that are representative of a site of interest. The user may select any of these options from the HRDB application main menu and follow a series of specification forms to select all the options necessary to complete the desired operation. The main menu (fig. 7) provides an interface for selecting each of these output options and an option for exiting the database application. The sequence of specification forms that are used to complete a desired operation is shown in figure 8. Although common forms are used for different options, the forms have customized features (such as titles and explanations) to cue the

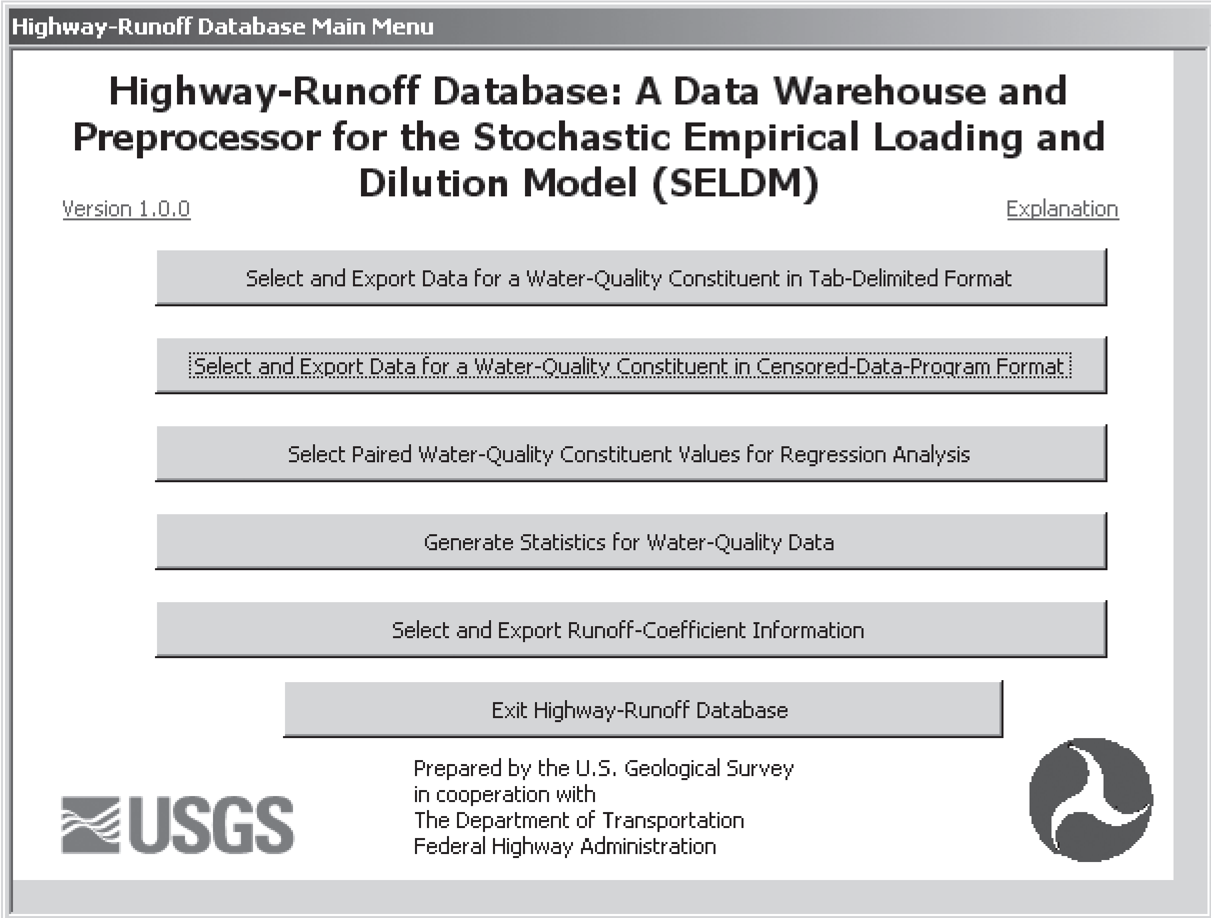


Figure 7. Main Menu Form for the highway-runoff database application.

user about the currently selected operation. For water-quality options, the user must (1) select the event type(s), (2) specify the constituent of interest, (3) select a data set or data sets that include the constituent of interest, and (4) select a monitoring site or sites with site characteristics—for example, the average daily traffic (ADT), location, impervious fraction, the presence of curbs, or the type of surrounding land use—that are similar to the characteristics of the site of interest. Once the user navigates through these common-use data-specification forms, the application forwards the user to the form that is specific to the individual task (fig. 8). Similarly, if the user chooses to export runoff-coefficient data, the user must select the event type(s), data set(s), and site(s) that have rainfall and runoff data and an estimated drainage area with the common-use data-specification forms (fig. 8). In each data-specification sequence, the user may either return to the previous form (by use of a “Go Back” button on each form) or exit the process and return to the main menu (by use of a “Quit” button). Use of the first four common-use data-specification forms is described here, and technical details about each main-menu selection and the resulting output are described in the following subsections.

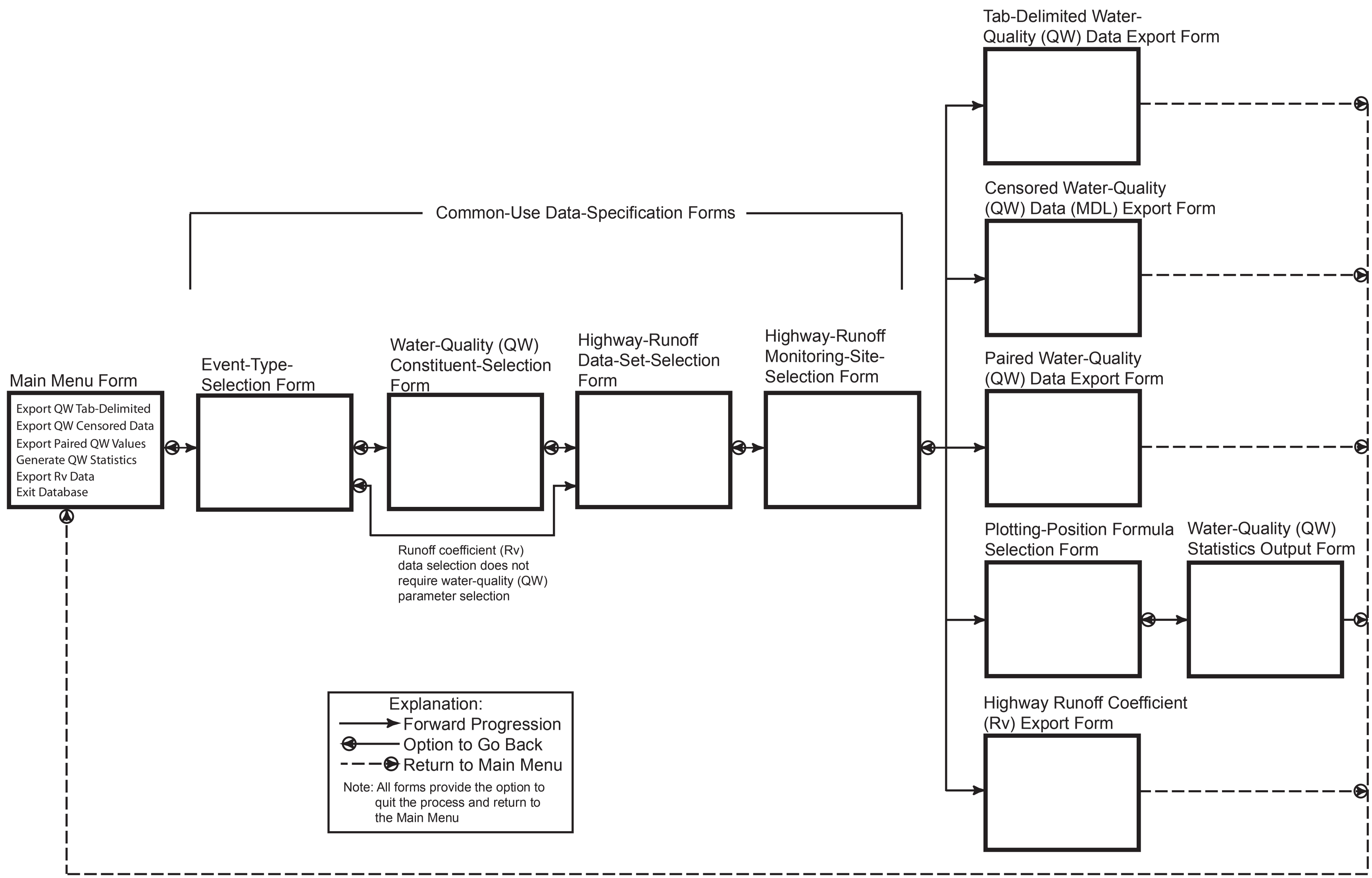


Figure 8. Highway-runoff database application schematic for processing water-quality (QW) and runoff-coefficient (Rv) data.

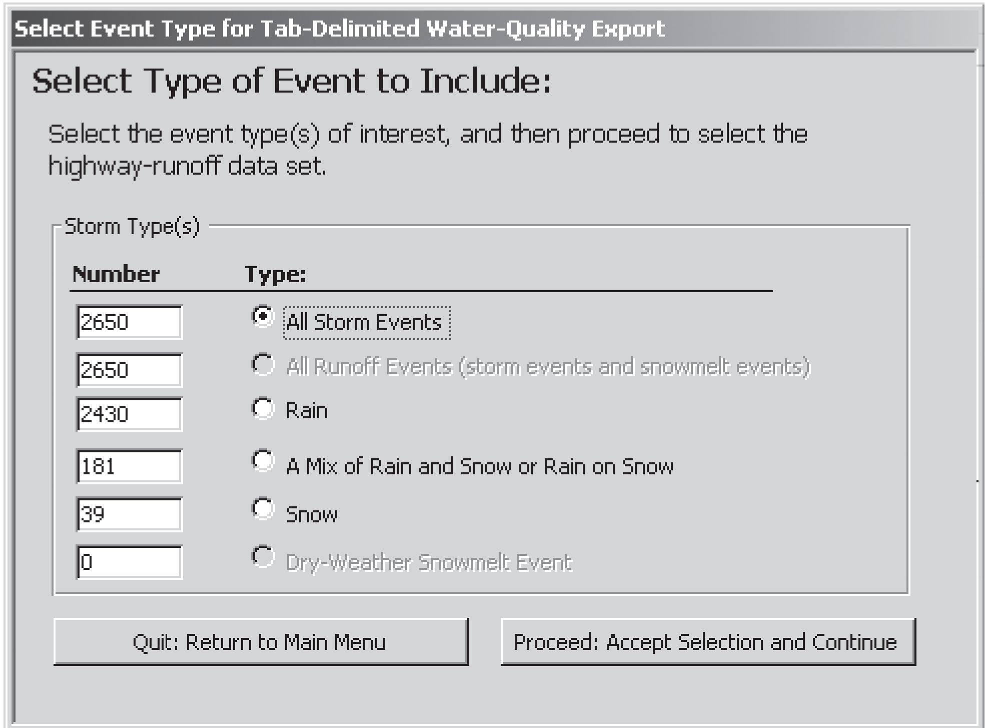
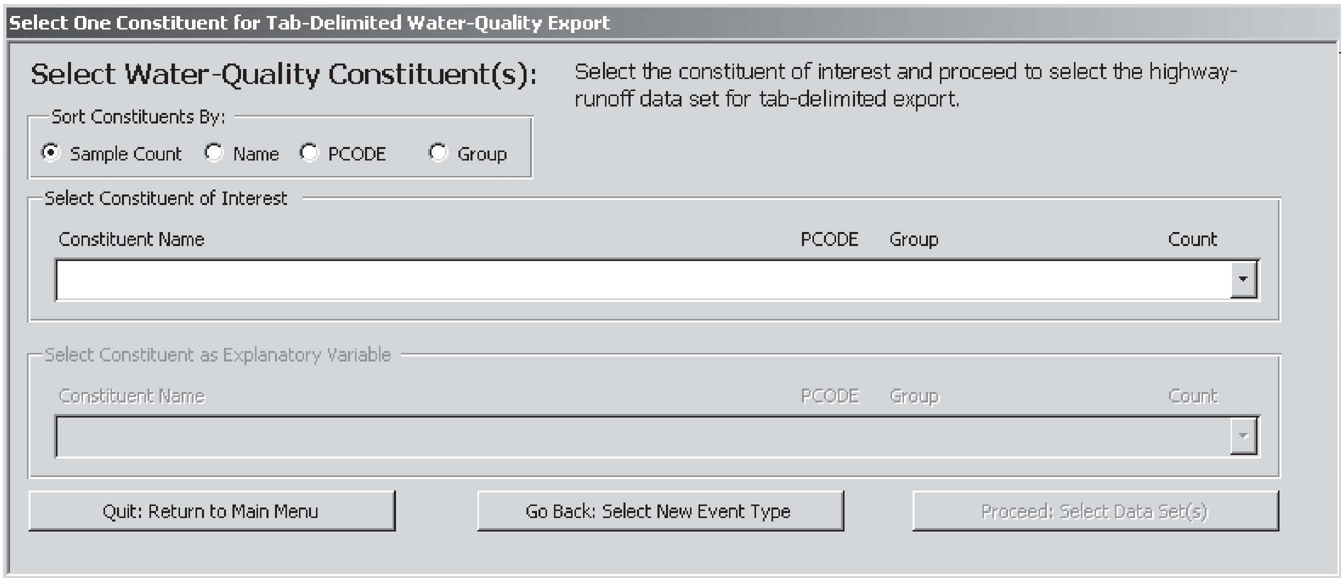


Figure 9. Event-type selection form for the highway-runoff database application.

The event-type selection form (fig. 9) is designed to allow the user to specify one or more event types to be used in the water-quality data or runoff-coefficient data selection process. Event types are specified because winter maintenance operations, such as sanding, salting, and plowing, may have a substantial effect on concentrations of a number of runoff constituents (Driscoll and others, 1990c). If winter maintenance operations increase constituent concentrations, a population of concentrations for all events may exhibit higher median and average values, greater variability, and an increased skew when compared to statistics for rain events. When the database application loads the event-type selection form, it runs several queries to count the number of storm (or runoff) events in the database. Storm events are defined as precipitation-runoff events. Storm event types include rain events, mixed events, and snow events (presumably with runoff). Mixed events are defined by Driscoll and others (1990c) as a mix of rain and snow or rain on preexisting snow. Runoff events include all storm events and dry-weather snowmelt events. The dry-weather snowmelt events are defined as runoff events that occur when air temperatures or solar radiation melt existing snow packs along a highway to cause measurable runoff flows. In the runoff-coefficient data-selection process, selections for dry-weather snowmelt events and all runoff events are disabled by the program because the dry-weather events are not associated with a specific storm-event precipitation volume. Currently (2006), there are no dry-weather snowmelt events recorded in the database.

The water-quality constituent selection form (fig. 10) is designed to allow the user to specify the water-quality constituent (or property) of interest. All water-quality constituents and properties in the database tables are organized by USEPA parameter code (PCODE). The PCODE is an unambiguous reference number that identifies the water-quality constituent or property, the sampling matrix, the sample type, and measurement unit (U.S. Environmental Protection Agency, 2005b; U.S. Geological Survey, 2005). There are concentration measurements for 116 water-quality constituents and properties in the HRDB and there are 7,427 possible water-quality constituents and properties identified by PCODE in the HRDB version 1.0. When the user selects any of the water-quality options on the main menu and selects the event type (fig. 8), the HRDB application queries the database to determine which water-quality constituents are included in the database for the selected event type(s), to count the number of values for each constituent, and to rank the constituents in descending order by the number of values available in the data set. By default, constituents are ranked by the number of samples available in the database in descending order so

A. Initial view of selection form



B. Active combo box on selection form

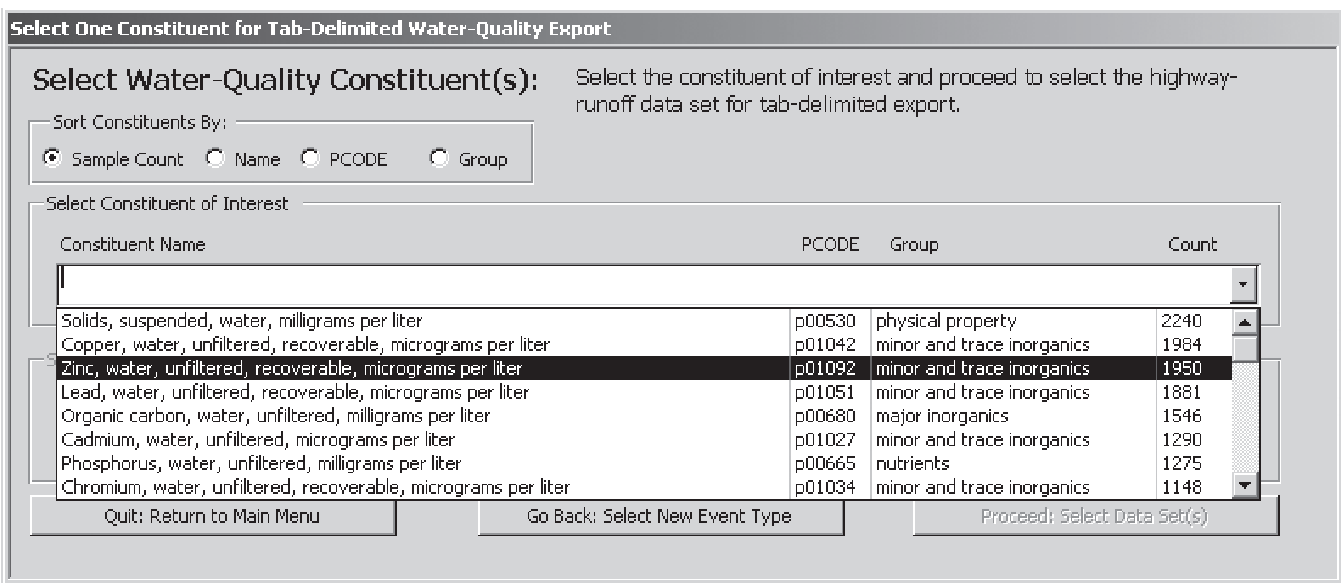
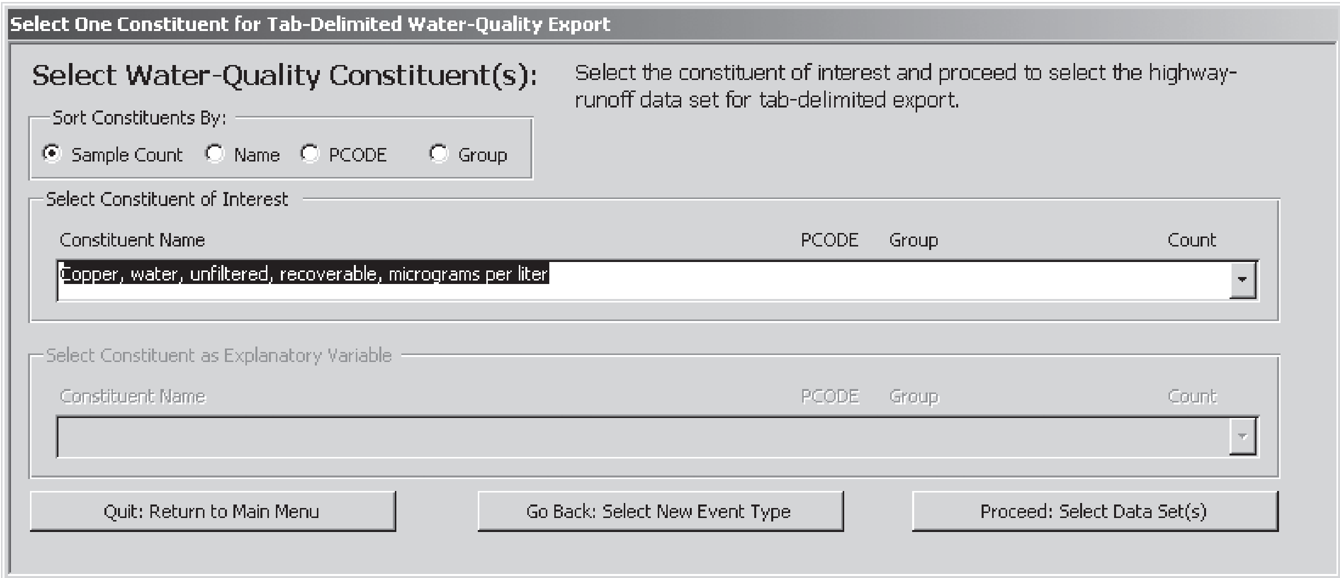


Figure 10. Water-quality constituent selection form showing *(A)*, initial view of selection form; *(B)*, active combo box on selection form; *(C)*, final view of selection form; and *(D)*, two-parameter selection form for the highway-runoff database application.

C. Final view of the selection form



D. Two-parameter version of the selection form

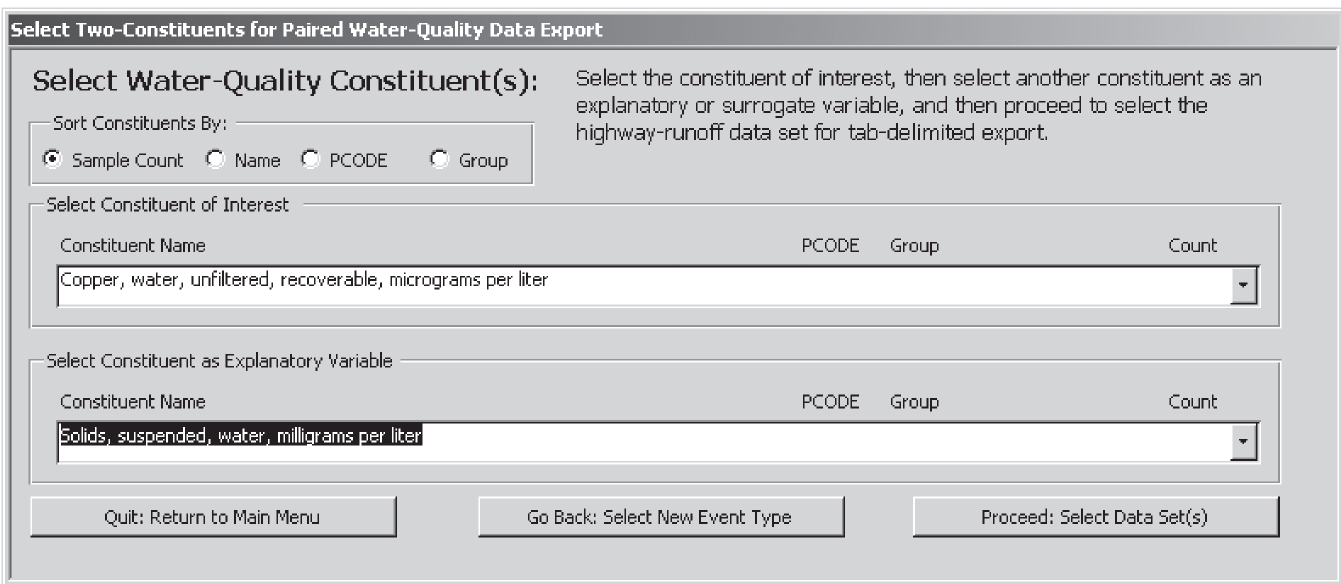


Figure 10. Water-quality constituent selection form showing *(A)*, initial view of selection form; *(B)*, active combo box on selection form; *(C)*, final view of selection form; and *(D)*, two-parameter selection form for the highway-runoff database application—Continued.

that the constituents of greatest interest for highway- and urban-runoff studies will be presented first in the selection list and the more obscure constituents with fewer analyses will be presented last. Most of the constituents of greatest interest for highway-runoff characterization (Granato, 2003) have more than 500 EMC samples in the database (fig. 5). However, the user also may choose to reorder the available constituents by name, PCODE, or parameter group by selecting the respective option (fig. 10).

When the water-quality constituent selection form appears (fig. 10*A*), the combo box(es) are blank, and the command button used to proceed to the next form is not activated. A combo box is a Microsoft form-control-object that can be used to select one object from a drop-down list of potential choices. Once the user clicks on the constituent-name combo box, a list of water-quality constituents including the name, PCODE, parameter group, and the number of EMC values in the database appears (fig. 10*B*). Once a constituent (fig. 10*C*) or, for the paired water-quality data option, constituents (fig. 10*D*) are selected, the name of that water-quality constituent appears in the combo-box window(s). The database application activates the “Proceed” command button once the water-quality selection(s) is(are) made. Constituent selection is the second step in each of the water-quality data-selection processes on the main menu because this choice limits subsequent selections to the data set(s) and data-collection sites with data for the event-type and constituent of interest. For example, all seven data sets include data for total copper, but only the Massachusetts data set (Smith, 2002) includes measurements of total cyanide.

The highway-runoff data-set selection form (fig. 11*A*) is designed with list boxes to allow the user to specify one or more highway-runoff data sets to be used in the water-quality data or runoff-coefficient data-selection process. A list box is a Microsoft form-control-object that can be used to select one or more objects from an on-screen list of potential choices. List boxes may have vertical and horizontal scroll bars that allow the user to view information that extents beyond the list-box dimensions. When the database application loads the data-set selection form, it runs a query to count the number of specified event type(s) and water-quality or runoff samples in the database by data set and populates the lower list box with a list of data sets that have the measurement(s) of interest. The lower list box includes the name of the data set, the range of sample-collection dates (period of record), and the number of samples of interest. The user may select a data set by left-clicking on the appropriate line in the lower list box. When this happens, a confirmation message appears in a pop-up message box. At this point the user may left-click “OK” to select the data set or “Cancel” to stop the selection. If a data set, is clicked and confirmed, the data-set name and period of record appear in the upper list box. To deselect a data set, the user must left-click the data-set name in the upper list box, and left-click the “Deselect Data Set” command button. If the user selects a data-set name in the upper list box and left-clicks the “Deselect Data Set” command button, a confirmation message appears in a pop-up message box. At this point the user may left-click “Yes” to deselect the data set or “No” to keep the selection.

The data-set selection form also provides a method to obtain bibliographical references for each data set. This form is provided because the citations allow the user to obtain and examine the source documents for the data in the database and to properly cite any data that are used. The need for such citations with water-quality databases has been identified by the NCHRP (Venner and others, 2004). If the user selects a data set and left-clicks the “Data-Set Reports” button, the Data-Set Citations form appears (fig. 11*B*). The Data-Set Citations form consists of an explanation, a large, scrollable text box and a close button. When the form opens, the citations for the selected data set are highlighted so that the user can easily copy and paste the citations from the text box into another computer application such as a text file, spreadsheet, or word-processing document. Left-clicking the “Close” button closes the form and returns the user to the data set selection form.

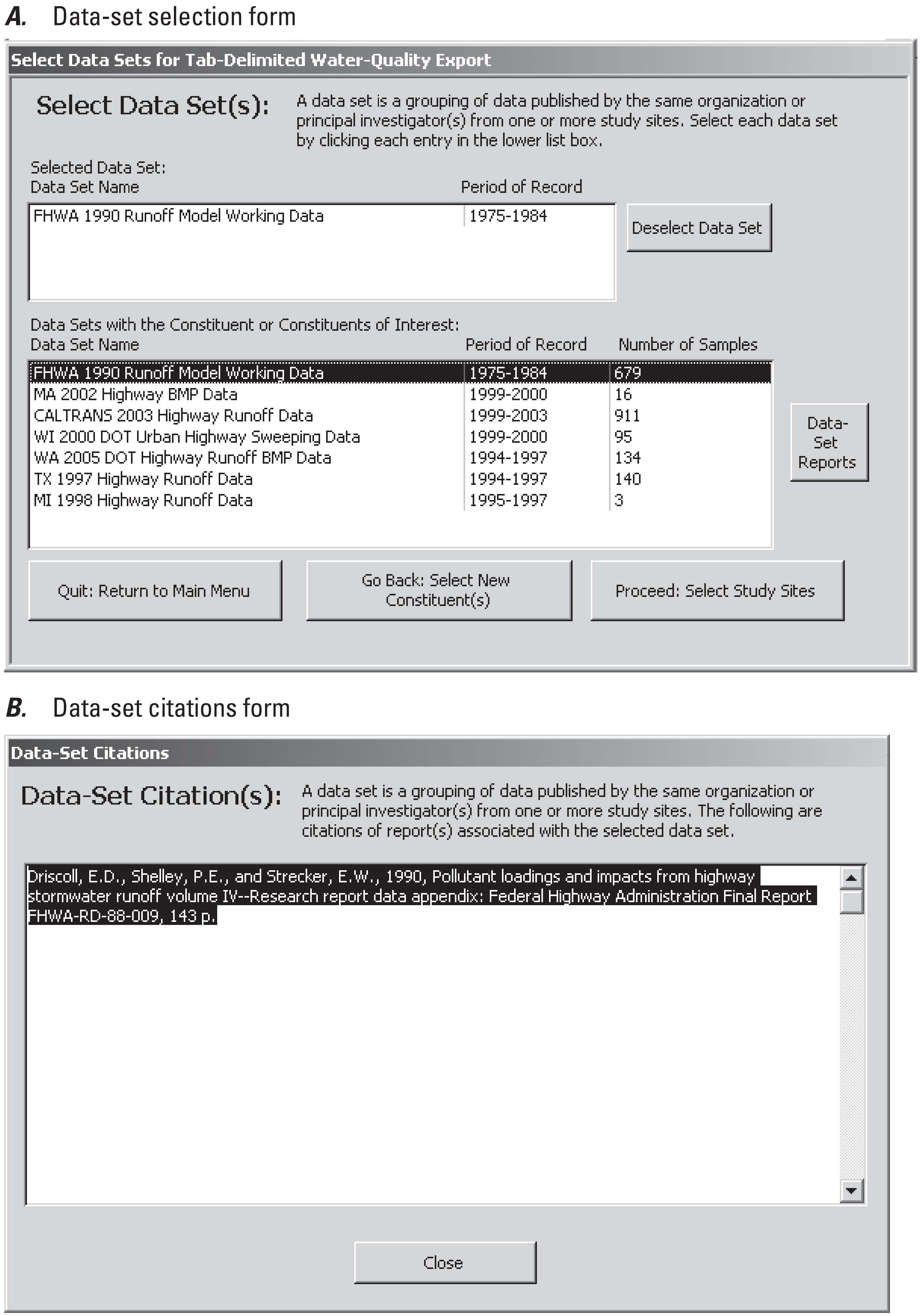


Figure 11. Data-set *(A)*, selection; and *(B)*, citation forms for the highway-runoff database application.

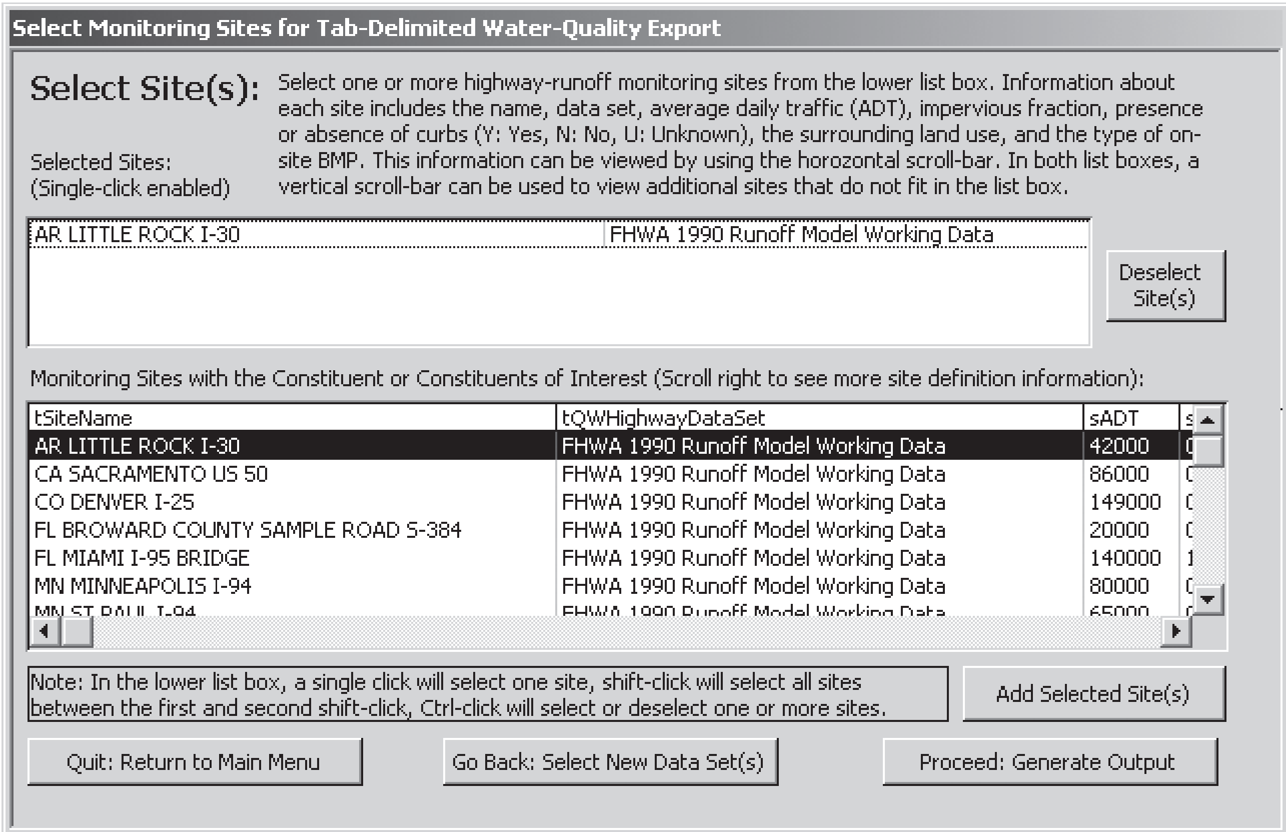


Figure 12. Site selection form for the highway-runoff database application.

The site-selection form (fig. 12) is designed with two list boxes and is used in a way that is similar to use of the data-set selection form. The user may select and deselect sites by clicking in the list boxes, clicking the appropriate command buttons, and responding to confirmation messages. The lower site-selection list box, however, is designed to allow selection of a single site (by clicking it), selection of two or more subsequent entries (by shift-clicking the first and last), and selection or deselection of multiple sites (by control-clicking individual sites). When the database application loads the site-selection form, it runs a query to count the number of specified water-quality or runoff measurements in the database by event type and data-collection site. The database application then populates the lower list box with a list of data-collection sites that have the measurement(s) of interest. The site-selection list box has a horizontal scroll bar that allows the user to view detailed site information such as name, data set, ADT, location, impervious fraction, the presence of curbs, the type of surrounding land use, the presence of upstream BMPs, and the number of water-quality or flow measurements for the parameter of interest. The headings for these columns are the database field-names, which are defined on the form and in the data dictionary on the CD-ROM accompanying this report. Once selections are made, they are added to the upper list box by clicking the “Add Selected Site(s)” command button. Sites may be deselected by right-clicking the site name in the upper list box and clicking the “Deselect Site(s)” command button.

This version of the HRDB application does not include a preprogrammed user interface for the tables that document sediment-quality data. This is because of the relatively small amount of sediment-quality data that are currently available and because of the technical complexities that must be considered by the user who may use sediment data to develop planning-level estimates of highway-runoff constituent concentrations. A user familiar with Microsoft Access and the highway-runoff data model could extract all necessary data by use of tables and user-defined queries in the database.

Select and Export a Water-Quality Constituent in Tab-Delimited Format

The first command button on the main menu (fig. 7) allows the user to select and export a water-quality constituent in tab-delimited format for use with other software applications including word processors, spreadsheets, and statistical packages. Once the user has selected the event type, water-quality constituent, the data set(s), and the site(s) (fig. 8), the application loads the Tab-Delimited Water-Quality Data Export Form (fig. 13). As the application loads the form, it runs a query to determine the total number of data points, the number of uncensored and censored data points, and the percentage of the data set that is censored and displays this information on the form (fig. 13). This export form allows the user to select a number of sort options and to segregate data by site. The user may choose to export explanatory information and data for each water-quality data point by selecting one or more export options on the form (fig. 13). Left-clicking the “Export Information” command button will activate a standard Microsoft Windows common-dialog box to allow the user to select the destination directory and file name for the tab-delimited data. The user may either “Go Back” to the previous form to select other sites or “Quit” to return to the main menu by left-clicking the appropriate command button.

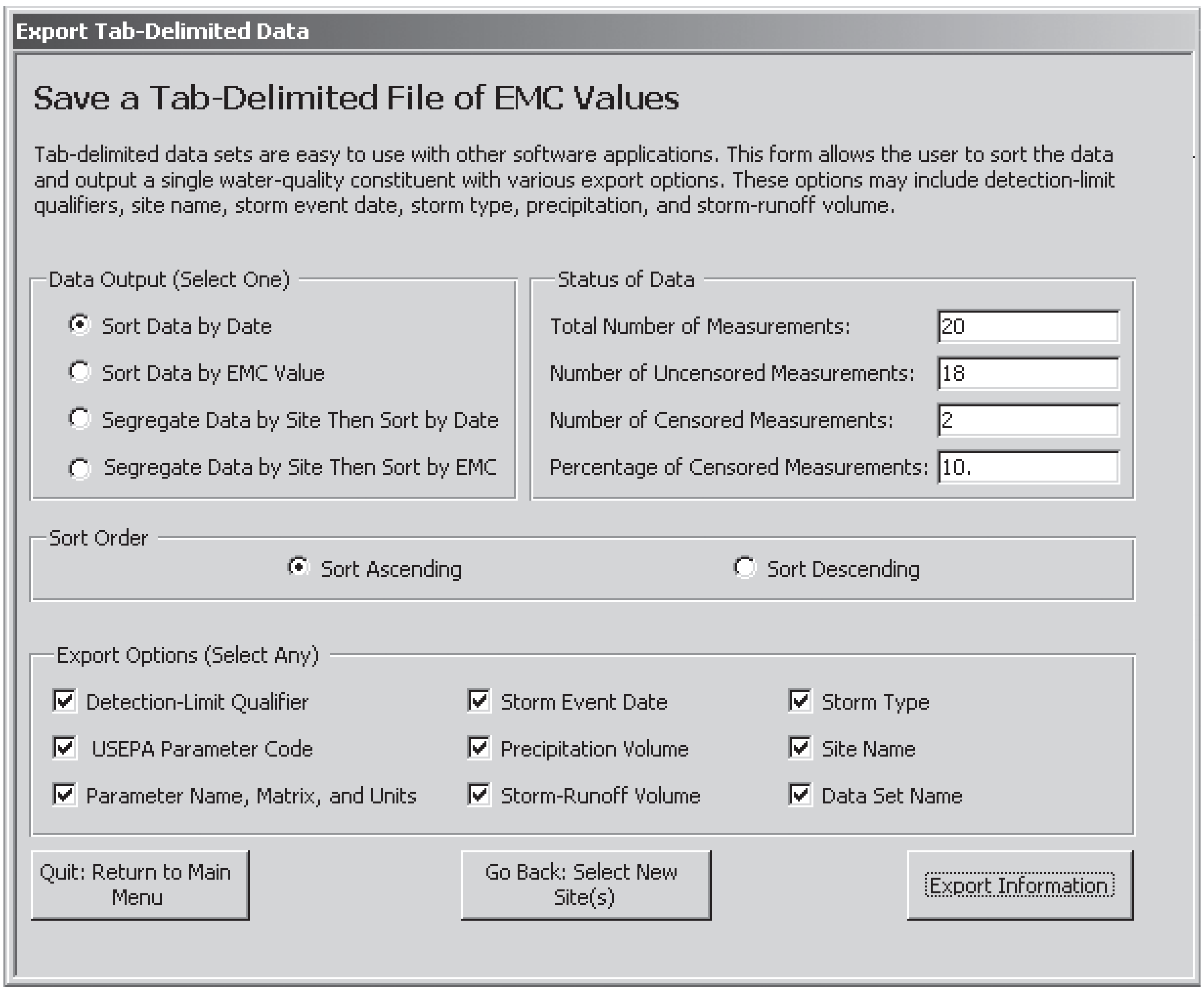


Figure 13. Tab-delimited water-quality data export form for the highway-runoff database application.

Select and Export a Water-Quality Constituent in a Format Suitable for Use with Computer Applications for Censored Data

The second choice on the main menu (fig. 7) allows the user to select and export a water-quality constituent in a comma-delimited format suitable for use with computer applications for calculating summary statistics for data with censored values, such as the USGS MDL program by Helsel and others (1988). The MDL program is an enhanced version of the original program for calculating summary statistics for data with values below (one or) multiple detection limits developed by the USGS (Helsel and Cohn, 1988). A version of the program compiled for Microsoft Windows 98 MDLWIN (Helsel and others, 1988) is available with example files and basic documentation on the CD-ROM accompanying this report and is available on-line (Helsel, 2004). The MDL program uses a robust version of the ROS method and the adjusted maximum likelihood (AML) procedure developed by Cohn (1988) to produce estimates of the arithmetic mean, standard deviation, median, and the 10th, 25th, 75th and 90th percentiles of the population of data.

The highway-runoff database produces the MDL input-file format described by Helsel and others (1988) as “File Format 2.” This format, which is also used by other detection-limit software, includes two comma-delimited entries: (1) the data or reporting limit and (2) the censored indicator for each data point. The censored indicator is coded 0 for censored data (below reporting limit), and 1 for uncensored data. Metadata about each sample (station name, sample date, and water-quality constituent name) also are output to the file in comma-delimited format. The MDLWIN program can accept up to 1,000 uncensored and 1,000 censored data points, but it requires at least 5 uncensored values to properly complete the calculations. This output-file format also is suitable for use with other software that is available for analysis of summary statistics for data with censored values (L.A. DeSimone, U.S. Geological Survey, written commun., 2005).

Once the user has selected the event type, water-quality constituent, the data set(s), and the site(s) (fig. 8), the application loads the Comma-Delimited Water-Quality Data Export Form (fig. 14). As the application loads the form, it runs a query to determine the total number of measurements, the number of uncensored and censored measurements, and the percentage of the data set that is censored. The application displays this information on the form. This export form allows the user to sort all values by EMC or to segregate by site and then sort by EMC. If the user is exporting multiple data sets by site, they can separate the sorted data in the database output file into MDL input files manually by use of a text processor such as NotePad, TextPad, or WordPad. Left-clicking the “Export Information” command button will activate a standard Microsoft Windows common-dialog box to allow the user to select the destination directory and file name for the comma-delimited data. The user may either “Go Back” to the previous form to select other sites or “Quit” to return to the main menu by left-clicking the appropriate command button.

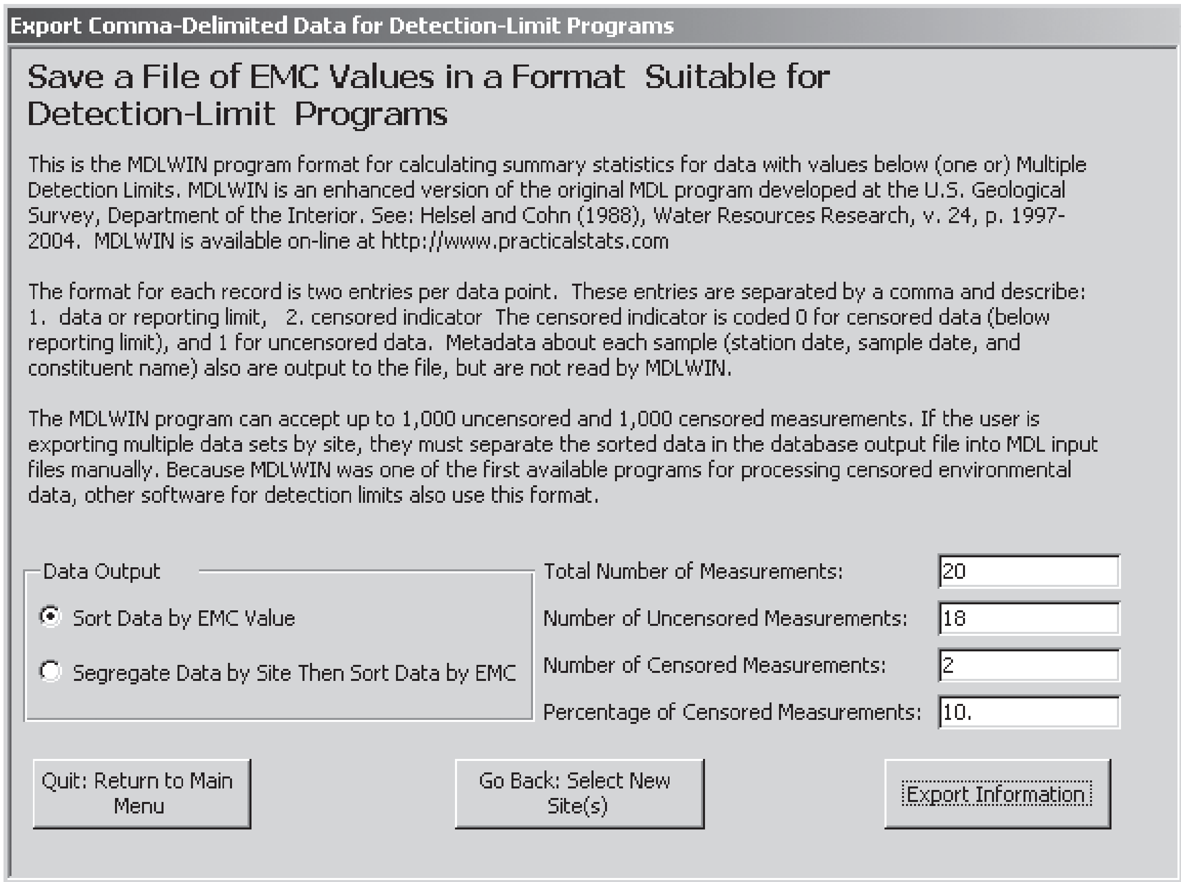


Figure 14. Export form for detection-limit programs for analysis of censored water-quality data for the highway-runoff database application.

Export Paired Water-Quality Data in Tab-Delimited Format

The third choice on the main menu (fig. 7) allows the user to export paired measurements of water-quality data in tab-delimited format for use with the Kendall-Theil Robust Line software (Granato, 2006). This format also is suitable for use with other software applications such as spreadsheets, commercial graphing packages, or statistical packages. Paired water-quality data may be used to examine relations between selected variables. Regression between variables may be used to estimate water-quality variables that are unavailable or are censored (Driscoll and others, 1990c; Thomson and others, 1996; 1997). If quantitative regression equations are identified, the user may estimate the values of water-quality constituents of interest from a surrogate variable. For example, the regression relation shown in figure 6 may be used to estimate SSC concentrations from TSS concentrations in highway runoff for use in calculating sediment concentrations in receiving waters downstream from a highway outfall. Trace metals and organic compounds are difficult and expensive to collect, process, and analyze properly and are commonly below detection limits in a proportion of filtered and whole-water samples (Breault and Granato, 2003; Lopes and Dionne, 2003). Regression equations may be used to estimate these constituents from SSC because trace metals and organic compounds commonly are associated with sediment in runoff and receiving waters. Finally, regression equations may be used for stochastic data generation, especially if the user wishes to maintain correlations between water-quality variables (Koch and Smillie, 1986; U.S. Army Corps of Engineers, 1993; Haan, 1994; Granato, 2006).

The paired water-quality data-file format has three tab-delimited columns. Each column in the output text file is identified by a header line in the first row that is the explanation for the data in that column. The first and second columns in the output text file include numerical data for use in regression analysis. The third column contains metadata about each sample in a semicolon-delimited string. The metadata column includes an "X:" and "Y:" designation for the first and second column, respectively. These designations are used to identify censored values with a qualification code (typically "<"). The metadata also includes the sample date, the site name, and the data set name for each XY pair in the selected data set.

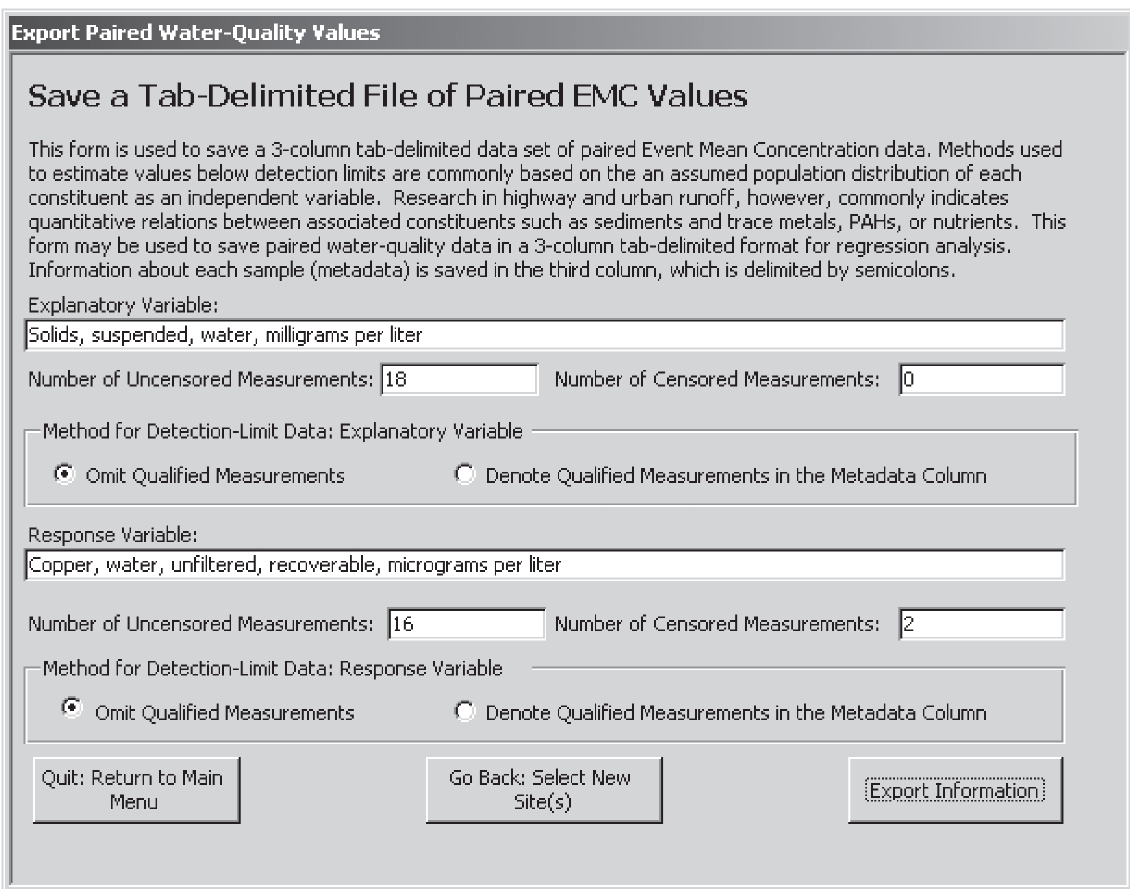


Figure 15. Paired water-quality data output form for the highway-runoff database application.

Once the user has selected the event type(s), two water-quality constituents, the data set(s), and the site(s) (fig. 8), the HRDB application loads the Paired Water-Quality Data Export Form (fig. 15). As the application loads the form, it runs a query to determine the total number of measurements, the number of uncensored measurements, and the number of censored measurements for each constituent. This information is displayed on the form. This export form has options for the user to include or omit censored values in both the explanatory and response-variable data columns. The default option is to omit these values because censored values may affect the regression equation. The option is provided so that the user may examine what values of the explanatory variable may be associated with censored values in the response variable. Left-clicking the "Export Information" command button will activate a standard Microsoft Windows common-dialog box to allow the user to select the destination directory and file name for this tab-delimited data. The user may left-click the "Go Back: Select New Site(s)" command button to move to the previous form and reselect the data-collection sites or left-click the "Quit: Return to Main Menu" command button to exit the paired-data export process.

Generate Statistics for Water-Quality Data

The fourth choice on the main menu (fig. 7) allows the user to select a water-quality constituent, generate statistics for the water-quality data, and export the results to a tab-delimited text file. The HRDB application calculates and outputs summary statistics of the retransformed values, the natural logarithm of the values, and the base-10 logarithm of the values independently. The summary statistics include the average, standard deviation, skew, and median. The statistics for each transformation are calculated separately because use of theoretical relations between summary statistics for different transformations may introduce bias in the statistical estimates. Bias may occur because the highway-runoff data sets for each site commonly have small sample sizes, and the logarithms of a sample of data may have nonzero skew coefficients (theoretically, the 95-percent confidence interval for the skew coefficient of a sample from a normal distribution is calculated as plus-or-minus two times the square root of 6 divided by the number of samples). If multiple sites are used to build a data set and the individual sites are not representative of one underlying lognormal distribution (for example representing highway runoff from large urban highways), the data may have a nonzero skew coefficient because it is a mixed lognormal distribution. If the data set does not include censored values, then the program calculates summary statistics using standard methods and provides the plotting position and lognormal Z-score of each EMC value.

If there are censored measurements and two or more uncensored measurements, the HRDB application will calculate summary statistics by use of the robust ROS method. A detailed description of the statistical and numerical methods used to calculate these summary statistics is contained in appendix 1. The resulting statistics, plotting position, and lognormal Z-score estimates are derived using the uncensored data and lognormally distributed estimates for each censored measurement. One value for each EMC measurement is provided so that the user may estimate different percentiles, but it should be noted that the individual censored-value estimates should not be treated as actual measurements when the user graphs the data or analyzes the data (Helsel, 2005). If the percentage of censored data is greater than or equal to 50 percent of the samples, the application will produce a censored median estimate from the ranked data (Helsel, 2005). A value of -9999 for any statistic indicates that there are not enough values to calculate the statistic.

If there are EMC measurements below one or more detection limits, the application also provides summary statistics for the uncensored data, estimates of population statistics by substituting the detection limit(s), one-half, one-tenth, one-hundredth, and one-thousandth of the detection limit(s). Substitution of zero for censored values is not included because it is assumed that highway-runoff EMCs commonly can be approximated by a lognormal distribution (Driscoll, 1990c; Thomson and others, 1996; 1997; Shumway and others, 2002). Use of statistics estimated from only the uncensored values or simple substitution methods are not recommended (Helsel and Hirsch, 2002; Helsel, 2005). Statistics from these methods are provided in the output to show the variability of estimates produced by substitution and to provide a range of mean and median values that are expected to bracket the true mean and median. The associated range of estimates of the standard deviation and skew, however, reflect the presence of detection limits rather than variability in the population of data.

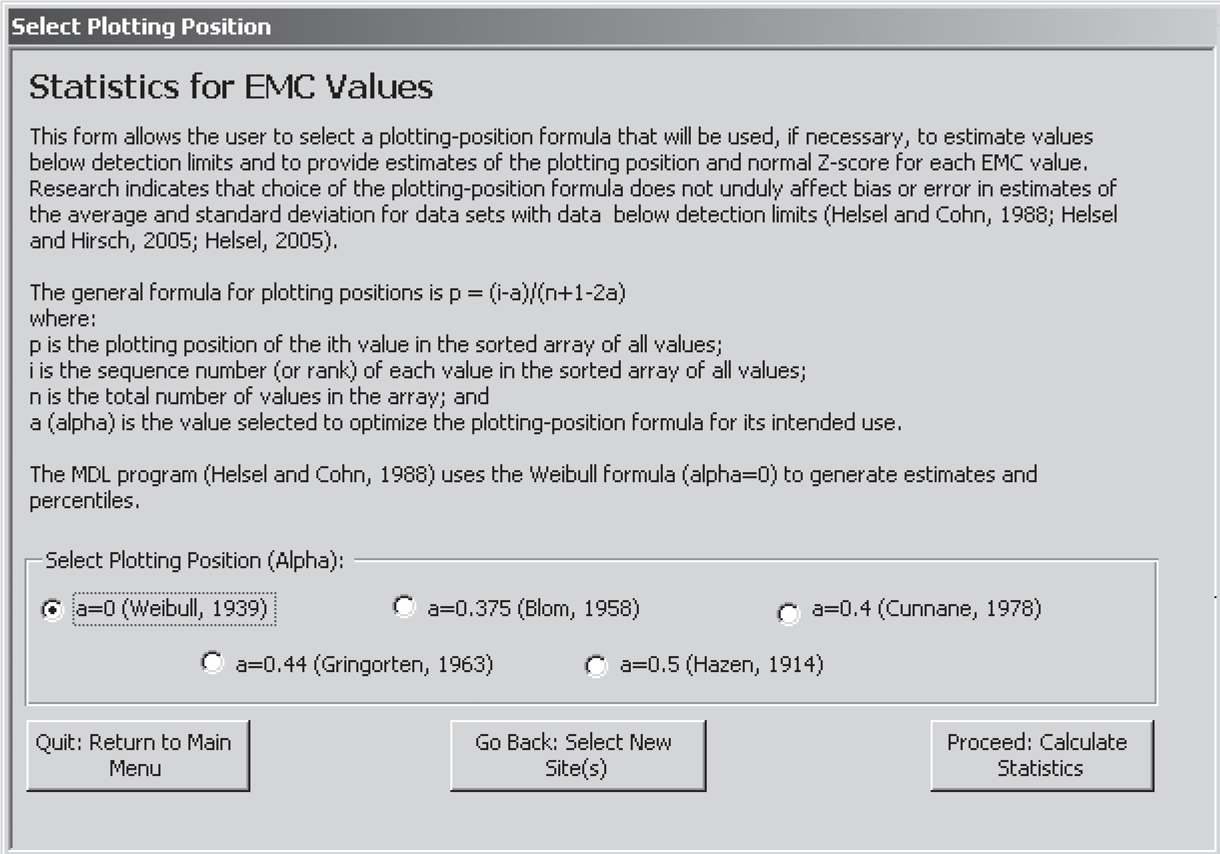


Figure 16. Plotting-position formula selection form for the highway-runoff database application.

Once the user has selected the event type(s), water-quality constituent, the data set(s), and the site(s) (fig. 8), the application loads the plotting-position formula selection form (fig. 16). The plotting position for each EMC measurement in a data set is the rank of the value (after being sorted in ascending or descending order by value) that has been normalized to a fraction between 0 and 1 by use of a plotting position formula. Theoretically, if the sample data set represents the underlying population of data, the plotting position represents the probability of each value in the data set. A number of plotting-position formulas have been proposed over the years, each having advantages and disadvantages for different populations of data (Helsel and Hirsch, 2002). The default selection in the database application is the Weibull (1939) plotting-position formula because this is the plotting position used by Helsel and Cohn (1988) to derive the nonparametric ROS method. Plotting-position formulas from Hazen (1914), Blom (1958), Gringorten (1963), and Cunnane (1978), also are available in the database application interface. Helsel and Cohn (1988) indicate that choice of plotting-position is of negligible importance for estimating the mean and standard deviation of the data. The Blom (1958) and Cunnane (1978) plotting-position formulas, however, are commonly considered to be preferable for (log) normal distributions (Helsel and Hirsch, 2002). Also, cursory examination of data in the highway-runoff database indicates that higher alpha values may decrease the value of skew calculated for the nonparametric ROS estimates. The choice of the plotting-position formula also affects the lognormal Z-score value associated with each plotting-position probability value. Each lognormal Z-score value indicates the distance of each value from the median as a fraction or multiple of the lognormal standard-deviation of the data.

When the user left-clicks the “Calculate Statistics” command button on the plotting-position formula selection form (fig. 16), the application loads the statistics form. As the

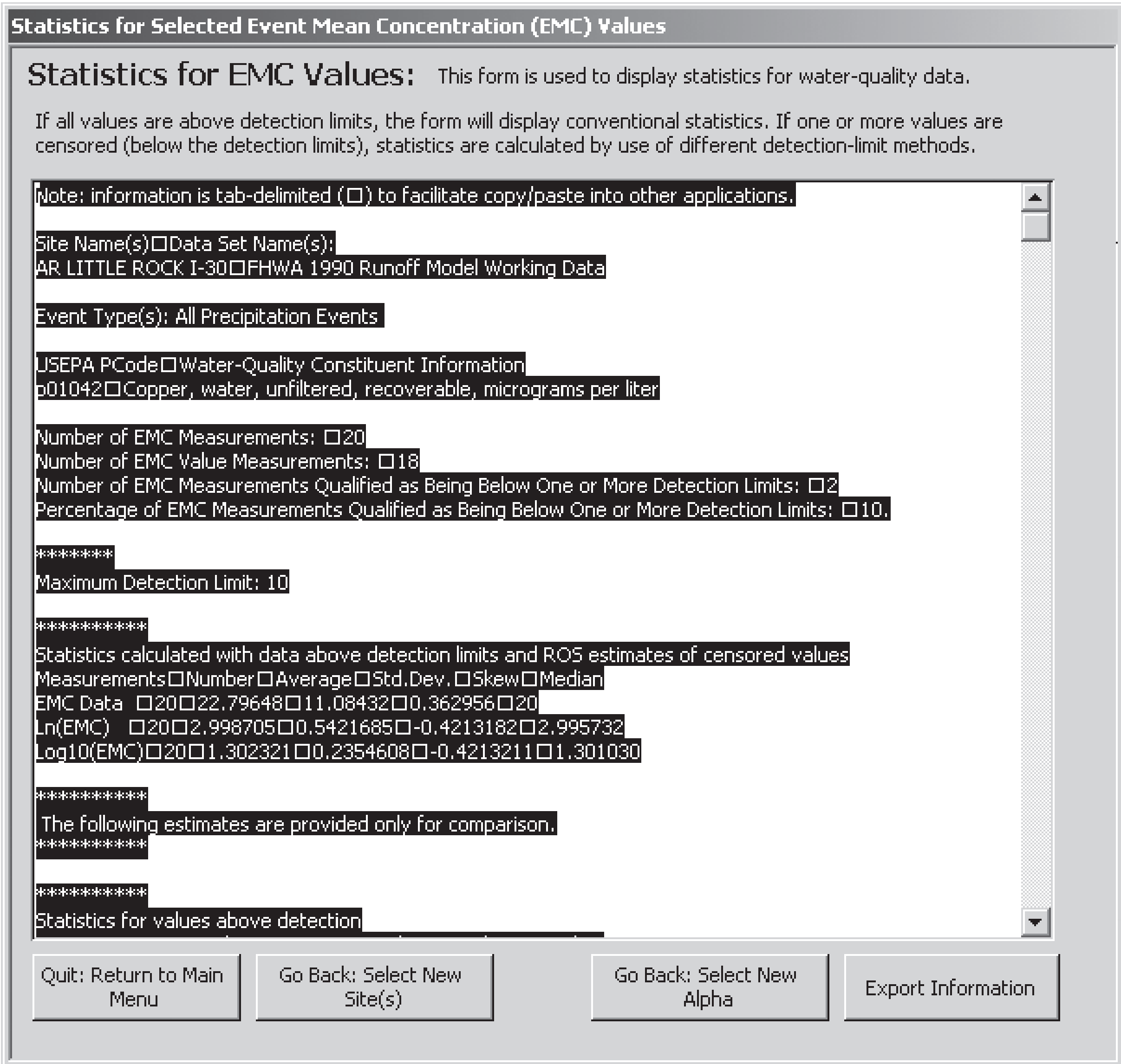


Figure 17. Statistical-estimate output form for the highway-runoff database application.

application loads the form, it runs a sequence of queries to define the data set, calculates the appropriate statistics, and populates a text box with the output values in tab-delimited format (fig. 17). The text box has a vertical scroll bar on the right-hand side so the user may scroll up and down to see all the statistical estimates. Left-clicking the “Go Back: Select New Alpha” command button allows the user to quickly change the plotting-position formula and see the effect on summary statistics. Left-clicking the “Export Information” command button will activate a standard Microsoft Windows common-dialog box to allow the user to select the destination directory and file name for the text file containing the statistical estimates. The tab-delimited format facilitates use of the results with many different software applications. The user also may highlight, copy, and paste results from the text box into another software application. The user may left-click the “Go Back: Select New Site(s)” command button to move to the site-selection form (fig. 12) and reselect the data-collection sites or left-click the “Quit: Return to Main Menu” command button to exit the water-quality statistics export process.

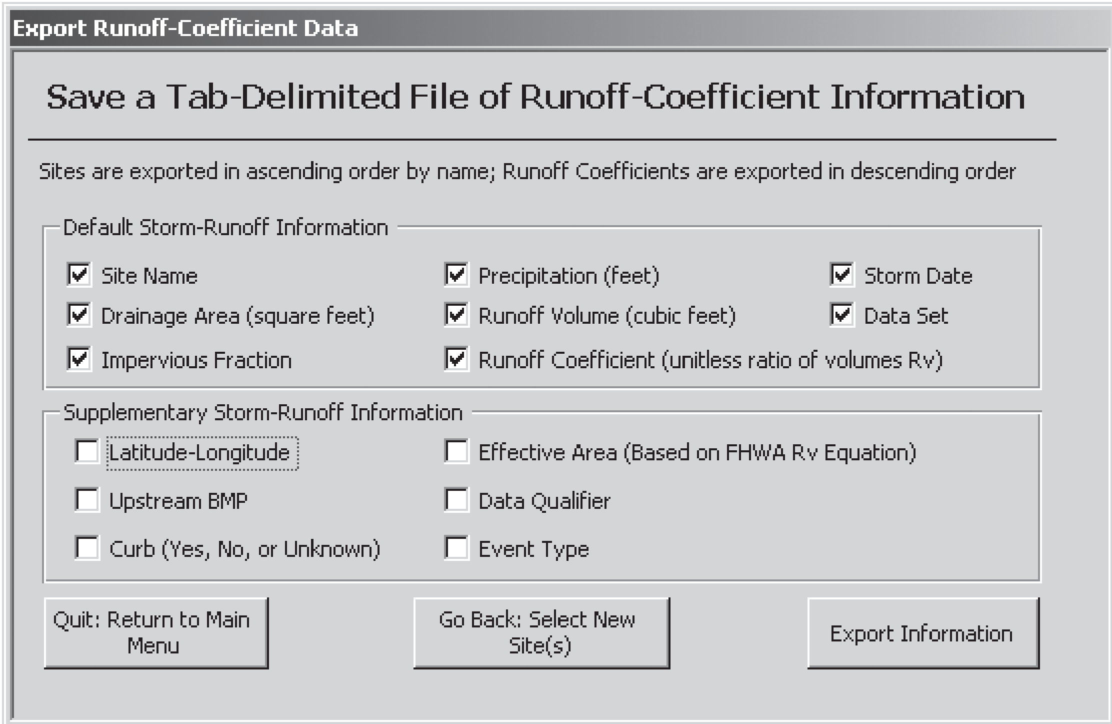


Figure 18. Runoff-coefficient information output form for the highway-runoff database application.

Export Highway-Runoff Coefficient Data in Tab-Delimited Format

The fifth choice on the main menu (fig. 7) allows the user to export highway-runoff coefficient data in tab-delimited format for use with other software applications such as spreadsheets, commercial graphing packages, or statistical packages. The export file includes eight default (mandatory) fields and six additional optional fields. The default fields include site name, drainage area (in square feet), impervious fraction, precipitation (in feet), runoff volume (in cubic feet), the dimensionless runoff coefficient, the storm date, and the name of the data set (fig. 18). The optional fields include the latitude and longitude coordinates, the presence of upstream BMPs (such as catch basins or swales) that may affect flows, and any data qualifiers from the storm table (tblStormEvent). In theory, all runoff coefficients should be less than or equal to 1, but many sites in the database have a substantial number of runoff coefficients greater than 1. This may be caused by use of a distant rain gage, by inaccuracies in measurement of precipitation and (or) runoff flow, and by inaccuracies in the basin-delineation process (Strecker and others, 2001; Church and others, 2003).

The effective drainage area is one of the optional fields that can be selected for output from this menu selection. The contributing drainage area to a given site may vary with instantaneous storm intensity if some percentage of runoff flow bypasses the drainage inlet. Small irregularities in the pavement can have a substantial effect on the effective drainage area to a monitoring site (Strecker and others, 2001). Vehicle speed and primary traffic direction can affect the amount of precipitation that is transported into or splashed out of a monitored catchment. The effective area of the catchment for each storm is calculated from the 1990 FHWA runoff coefficient regression equation (Driscoll and others, 1990c) to allow the user to examine the veracity of estimates of precipitation, runoff, and drainage area in terms of physical site characteristics. If the average of these estimates is used to characterize a given site, this will center the storm-to-storm variability in the site-specific runoff coefficients on the 1990 FHWA regression equation estimate. The effective area of the catchment for each storm is

$$Effective\ Area_{square\ feet} = \frac{Runoff\ Volume_{\ cubic\ feet}}{(0.7 * Impervious\ Fraction + 0.1) * Precipitation_{\ feet}} \quad . \qquad (1)$$

Once the user has selected the event type(s), data set(s) and the site(s) (fig. 8), the application loads the Highway-Runoff Coefficient Export Form (fig. 18). This export form allows the user to choose to include or omit the optional storm-runoff information. Left-clicking the “Export Information” command button will activate a standard Microsoft Windows common-dialog box to allow the user to select the destination directory and file name for the tab-delimited data. When the user selects the export command button, the application runs a query to convert all units to the basis of feet, calculates the runoff coefficients and effective areas, and prints this information to the output file. The user may left-click the “Go Back: Select New Site(s)” command button to move to the previous form and reselect the data-collection sites or left-click the “Quit: Return to Main Menu” command button to exit the runoff-coefficient export process.

Qualification Code Maintenance Form

The Qualification Code Maintenance Form is not on the database main menu, but may be activated from the Microsoft Access database-forms interface. The database is designed such that uncensored measurements will have a null qualification code, and censored measurements will have a text value (typically “<”). The database application is designed with the assumption that censored values are below detection limits (left-censored data). Values that are greater than quantification limits (right-censored data) are relatively rare. Therefore, right-censored data should be identified with the greater-than symbol (>) in the EMC value comment field. Activation of the Qualification Code Maintenance Form launches a series of queries that provide a count of each type of value in the qualification-code field (fig. 19). Program code for this form also retrieves and counts any nonnumeric or null EMC values. The results of this search also are presented in a message box. If null values are present, these values are printed in the text box on the Qualification Code Maintenance Form (fig. 20). If there are qualification codes that are not null or composed of one or more spaces, left-clicking the “Fix Blank Codes” button will nullify these fields. Left-clicking the “Quit: Return to Main Menu” command button will close the form and return the user to the main menu.

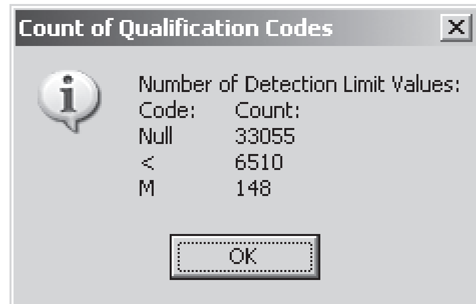


Figure 19. The qualification-code message box for the highway-runoff database application.

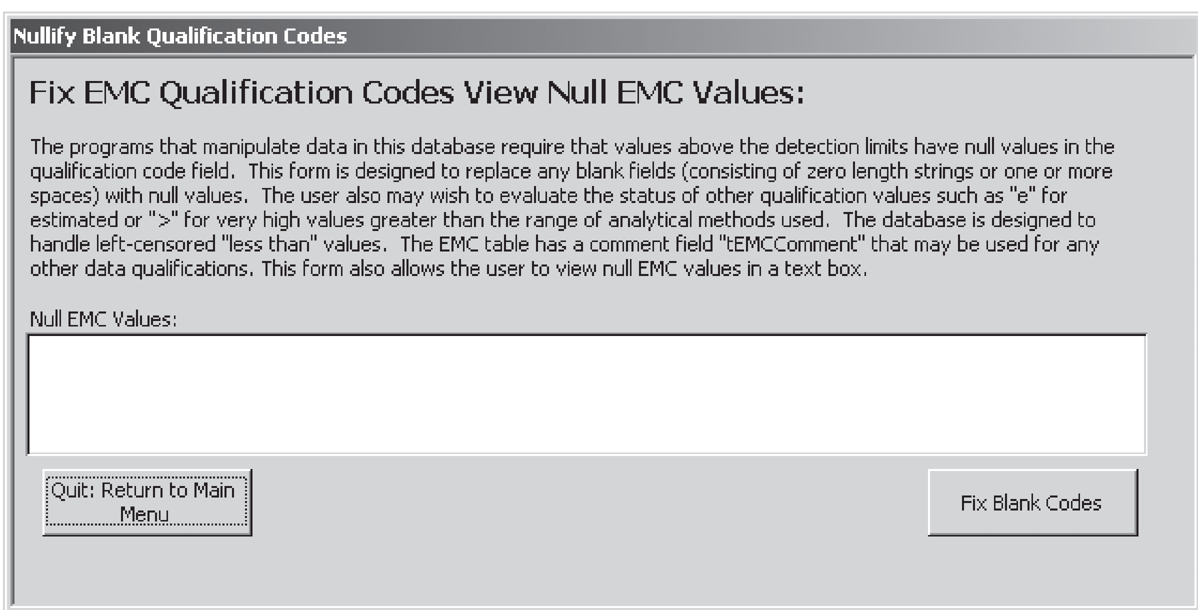


Figure 20. Qualification code maintenance form for the highway-runoff database application.

Database Design

The highway-runoff database was designed and implemented following standard relational database design and documentation methods (Fleming and von Halle, 1989; National Institute of Standards and Technology, 1993; Roman, 1997; Granato and Tessler, 2001; Granato, 2004). Each table in the relational database is designed to characterize a data element in the database. Examples of unique data elements include highway-site monitoring stations, storm events, and EMC measurements. One or more tables are created for each unique data element. Tables consist of one or more fields (columns) that define the characteristics of each attribute of the data element. An entity is defined herein as one member of a data element. For example, one highway site may be an entity defined in a table describing highway sites. Each unique entity (commonly referred to as an entity instance) in a table is defined within a record (row). The data in each field (column) in the record (row) documents one of the unique characteristics of the entity instance. In the relational database design, each row is a unique record because each row must have a combination of field values that define a unique entity instance. For example, a table may be used in a relational database to provide a standard list of site pavement types. At a minimum, a table characterizing the pavement type should contain two fields, an index number and a descriptive name for the pavement type.

The power of a relational database lies in the ability to compartmentalize each unique data entity in a highway-runoff data set into one or more individual tables that characterize the data entity without duplicating or losing the information that describes the relationship between individual entities. For example, the highway-runoff data are organized by data set (data from a group of sites, which are related to a published study), by highway site, and by storm event. Each data set may have data from one or more sites; each site may have data from one or more storms; and each storm may have data for one or more water-quality constituents. It would not be efficient to repeat the data set, site, or storm information with each of the 39,713 water-quality constituent measurements. For example, there are (on average) about 15 sites per data set, 25 storms per site, and EMC measurements for 15 constituents per storm (fig. 1). Similarly, each of the 39,713 water-quality constituent or property measurements (in table tblEMCValues) is defined by a parameter identification number (Parameter_ID) that serves as a relational link to a separate table (tdsUSEPAParameterCodes) containing this identification number, the USEPA 5-digit parameter code, a description of the parameter group, and the full text description of the parameter (property or constituent). In this way the detailed parameter description, which may require more than 100 text characters, is listed once (in table tdsUSEPAParameterCodes) and only the 1- to 4-digit identification number (Parameter_ID) is repeated with each of the 39,713 EMC measurements (in table tblEMCValues).

Table- and Field-Naming Conventions

Naming conventions are necessary to communicate the identity and contents of the database object unambiguously. Consistent use of a standard naming convention facilitates an understanding of design elements and relationships in the design of the database. A standard naming convention also is an effective documentation tool in the development and use of the database because the user can interpret the purpose and scope of each database object by examining its name.

Table names in the highway-runoff database consist of a three-letter functional prefix and a definitive table name. Tables beginning with the prefix “tas” (table, association, simple) or “tad” (table, association, data) are association tables that link information in two or more data tables by use of the primary key-fields from each table. The “tad” table is designated as such because it also contains one or more additional data fields that provide data associated with the relationship between entities in the parent tables. Tables beginning with the prefix “tbl” are data tables, which characterize individual data entities. Tables beginning with the prefix “tdm” (table, domain, multiuse), “tds” (table, domain, static), or “tdx” (table, domain, extendable) are domain tables, which are used to provide standard choices to characterize data elements. The “tdm” tables do not have a numeric key-field and so may be used repeatedly in a table to supply a drop-down list of standard choices. The “tds” tables contain fixed standard choices indexed by a numeric key-field. The “tdx” tables contain standard choices indexed by a numeric key-field, but the contents of these tables may be extended by the user. Tables beginning with the prefix “ttbl” are temporary-data tables that are used by the database application to temporarily store the results of calculations or to facilitate data manipulation.

The HRDB application also contains graphical-interface forms queries, and Visual Basic code modules. Form names begin with a “frm” prefix, query names begin with a “qry” prefix, and module names begin with a “mod” prefix. Use of these prefixes facilitates identification of different components within the database and in the associated Visual Basic code.

Field-naming conventions are based on field type. Field names throughout the database are based on whole words (such as “tLocationDescription”) or well-recognized abbreviations (such as “dLatLongAccuracy”) that are capitalized to emphasize the individual words. Field names for indexed database-key fields, which are all long integers, have the format Name_ID (such as “Site_ID”). This key-name convention maintains compatibility with the USGS-

FHWA NDAMS database design (Granato and Tessler, 2001; Granato, Dionne, Tana, and King, 2003). Most other fields start with a lower-case prefix letter that indicates the data type of the respective field. These one-letter prefixes are “b” for Boolean (which is a numeric field using -1 and 0 for yes and no, respectively), “d” for double-precision real values, “s” for single-precision real values, “i” for integers, “l” for long integers, “m” for memo fields, and “t” for text fields. Date-time field names are preceded by the three-letter lower-case prefix “dtm.” Detailed descriptions of field types are shown on plate 1. The lower-case prefix conventions for fields within some domain tables imported from the NDAMS database are not followed (for example, the field “State” in table “tdxState”) to preserve the backward compatibility in the design of these tables.

Table- and Field-Definition Conventions

Table and field definitions provide descriptive information about each table and each field within a table. Table definitions are entered in the description-property window accessed by right-clicking the table and choosing the properties setting on the pop-up menu. The table definition is available to the database user in the table-object window of the Microsoft Access interface. The table definition may be retrieved by use of a macro or a Visual Basic module when a full Microsoft Access application is developed from the database design. Similarly, a definition for each field within each table is entered in the table-design window. Once field definitions are entered, the definitions are available to the database user in the information bar in the lower left of the Microsoft Access interface. Microsoft Access automatically links these definitions when the table fields are used in Microsoft Access queries or forms. The use of table and field definitions provides necessary metadata about each object for use or development of the database. The conventions used for table and field definitions are not as rigorous as the table- and field-naming conventions, but the definitions are implemented systematically.

Entity/Relationship Diagramming Conventions

Entity/Relation (E/R) diagrams are used to visualize database designs. Several different display and notation methods are in common use for E/R diagrams, but all share similar characteristics. The Information Engineering (IE) relation notation and style (National Institute of Standards and Technology, 1993), enhanced by the database naming conventions, are used to document the database design. An E/R diagram that illustrates several diagramming conventions is shown in figure 21. In this E/R diagram, boxes are used to denote an entity, which is a single table in the physical database. Each entity box has its name at the top. Within the box are one or more entity attributes, which are fields in the physical database. Field types in figure 21, and other E/R diagrams in the text, are generalized as date-time, number, or string. With the exception of key fields, the data type is indicated by the field-name prefix. Field types are more fully defined on plate 1 and in the data dictionary on the CD-ROM accompanying this report. Connecting lines represent the defined relationship between entities.

In the E/R diagram, the primary key (PK) for each entity (composed of one or more attributes) is listed at the top of the attribute list within the entity box and is separated from the other attributes by a horizontal line. When a PK from one table (the parent) is passed to another table (the child) through a relation, the corresponding foreign key (FK) in the child table is designated FK in the diagram. If the FK is part of the PK in the child table, the relation is said to be strong and the relation line is solid in the diagram (this will always be true for association tables). If the FK is an attribute of a child table, the relation is said to be weak, and the line is dashed. To further help the user visualize table dependencies, tables in a strong relationship are shown with rounded corners (for example, tasHighwayDataSetCitations; fig. 21), whereas tables that do not have FK dependencies in their PK are shown with squared corners (for example tblQWHighwayDataSet; fig. 21). Key fields, designated as Name_ID, are all long integers. Key fields also are generalized as numbers on the E/R diagram figures in the text. However, key fields are identified as AutoNumbers in the parent tables that use the Microsoft Access AutoNumber feature on plate 1 and in the data dictionary. The AutoNumber feature is a utility that will produce sequential (or random) long-integer values to generate key values. The corresponding values are designated as long integers where they appear as FK values in the child table.

Each relationship line has a direction and cardinality. The direction is recognized by the origin end (parent entity in the relation), which either does not have a symbol or has an open diamond (when the relationship provides an optional FK value), whereas the target end (child entity in the relation) has a filled circle (dot). These relation symbols, however, do not define cardinality of the relations. Cardinality defines how each record in the origin entity (parent) relates to records in the target entity (child). Relationships between entities may be defined as one-to-one (1:1) or one-to-many (1:n). Each of these relationship types also may include a one to zero (1:0) if there is a parent record without entries in the child table. In a 1:1 relationship, each entity instance in the parent table may have zero or one match in the child table. For example, the table tblQWHighwayDataSetQAQC has a 1:1 relationship with table tblQWHighwayDataSet (fig. 21). In theory, tables with 1:1 relationships could be merged in a fully normalized database. This type of relation is used when there is an operational or administrative reason for segregating data tables. In a 1:n relationship, each entity instance in the parent entity may have zero, one, or more matching instances in the child entity. For example, the table tasHighwayDataSetCitations has a 1:n relationship with table tblQWHighwayDataSet (fig. 21).

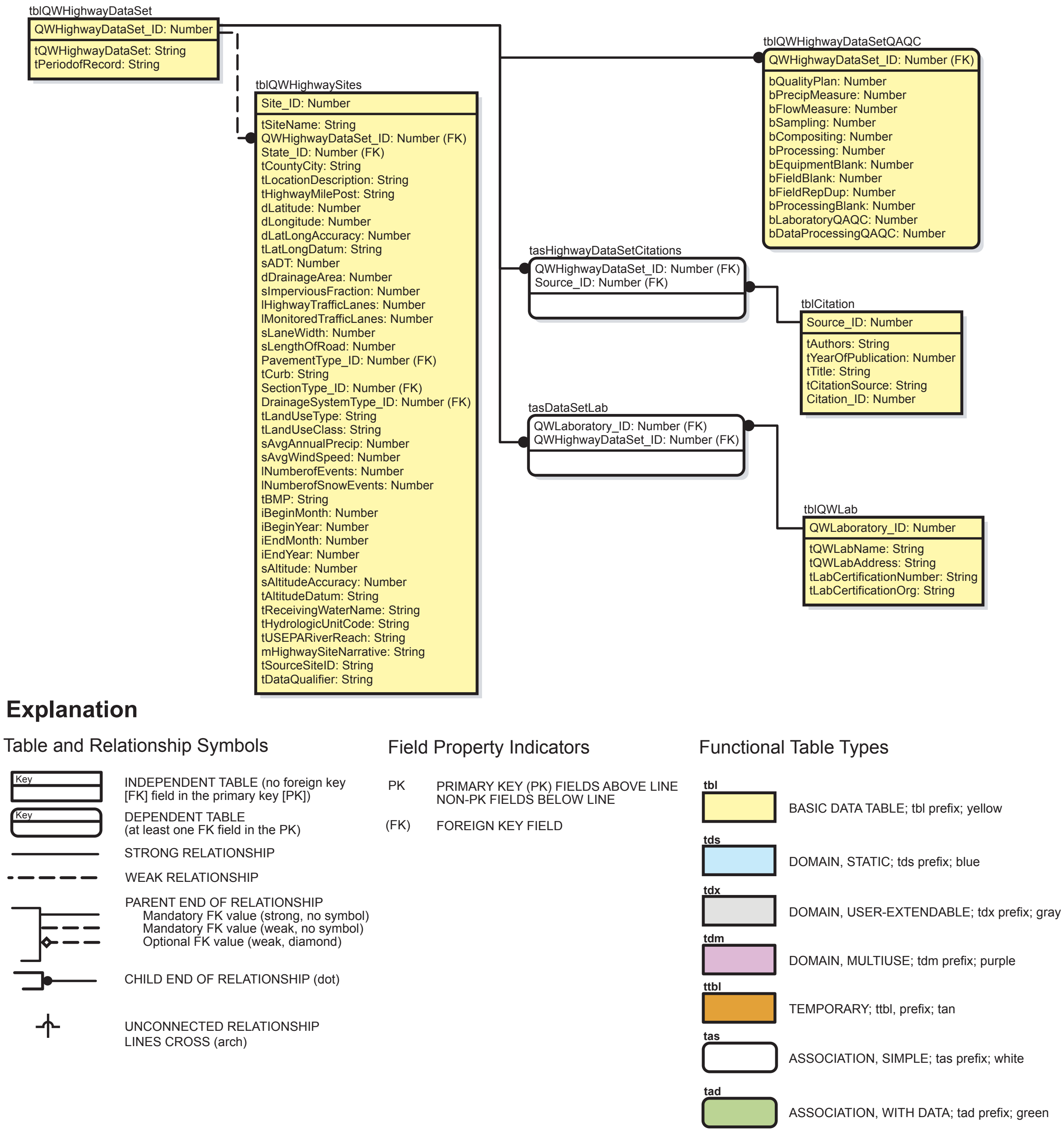


Figure 21. An entity-relationship (E/R) diagram showing a graphical representation of tables, fields, and relationships of the data structure for the highway-runoff data sets.

The cardinality of a relationship is implicit in the database diagram. Weak relationship lines (dashed lines) define one-to-many (1:n) relationships because a FK may be repeated in the child table. Strong relationship lines (solid lines) define a one-to-many (1:n) relationship if there are two or more PK fields in the child table. Strong relationship lines (solid lines) define a one-to-one (1:1) relationship if there is one shared PK field in the dependent child table.

Almost all relationships in the highway-runoff database are one-to-many (1:n). This type of relationship is used when each parent can have none, one, or more than one child, and each child must have a parent (the FK cannot be null). For example, in figure 21, the table tblQWHighwayDataSet has a parent-to-child relationship with tblQWHighwaySites. Each QWDataSet_ID record can be used to classify zero, one, or more than one highway-monitoring sites, and each highway-monitoring site must be attributed to one highway-runoff data set. The table tblQWHighwaySites, however, is an independent table because a unique Site_ID identifies each monitoring site. A one-to-zero relationship allows the user to create a record in a table that has no relationships in other tables. This type of relationship allows the user to populate a domain table with a list of all permissible values before other data are entered into a database. These values are not used until they are needed. For example, the domain table tdsUSEPAParameterCodes contains 7,427 standard USEPA water-quality constituent parameter codes. Currently, only 116 of these parameters are used, but the others are available for use as the database is populated.

Database Design Documentation

Design documentation facilitates current use and potential modification of the database for future use. The design of the database and implementation of the database application is fully documented on the enclosed CD-ROM in four ways, including

- documentation in the structure of the highway-runoff database,
- a data-dictionary file,
- a detailed database-design diagrams, and
- this report.

The highway-runoff database contains four types of design documentation in the Microsoft Access file: table definitions, table-design details, database-relationship information, and open-source application code in the forms and modules in the database. Microsoft Access is an object-oriented application. Each object (such as a field, table, relationship, or query) has standard properties that are documented. For example, each object has a description property that is used in the database design to describe the purpose and scope of each object. The documentation in Microsoft Access is useful for examining individual objects, but not for providing an overview or for illustrating the overall design of the database. Therefore, the data dictionary, database-design diagrams, and this report are provided to meet this documentation need.

Table names and descriptions identify the purpose and scope of each table. Table names and descriptions are visible in the table-object window when the “Details” view is selected. Alternatively, table definitions may be viewed and edited by right-clicking a table and choosing the “Properties” option, which activates the table-properties window.

Each table is composed of fields, which have names, descriptions, and other properties. Field names and descriptions identify the purpose and scope of each field. Each field description is visible in the status box at the lower left of the Microsoft Access interface screen when the table is open in datasheet view and the field is active. Properties of each field, including

- the presence of keys as denoted by a key symbol,
- field names,
- data types,
- description, and
- specific field properties

are visible in the design view of each table. The table-design view allows the user to assess and manipulate field properties. The table-design view should be used carefully because changes in field properties may corrupt the database and its contents.

The relations between database tables may be viewed by using the tools menu and selecting the relationships option, which activates the Relationships window. The highway runoff database, however, is complex enough to limit the clarity of information available in this view if the entire database is viewed at once. To view individual subject areas, users may activate the Microsoft Access relationships window, add the table(s) of interest with the “Show Table Button,” and then click the “Show Direct Relationships” button on the tool bar to see all tables that have relationships with the table of interest.

The visual basic for applications (VBA) program code that is used to respond to user input, manage data, and calculate statistics is fully documented in the forms and modules of the database. To view the code behind each form, the user may open a form in design view and left-click the code icon. This will activate the Microsoft VBA interface. The highway-runoff database is dependent on two VBA modules in the database. The module modPublicVariables contains public variables, which are used throughout the application code, and generic subroutines. The module modPublicStats contains the statistical subroutines that summarize user-selected data sets for output. The user should be cautioned that changes in this code, in table or field names, or in the names of form controls in the database may disable the application.

A computerized data-dictionary file is provided to facilitate examination of the database design and to document the completed database. Complete documentation of the table names, table descriptions, and information about each of the fields in the database is provided in the data dictionary file. The data-dictionary is an Adobe PDF file on the CD-ROM named HRDDv01.pdf. This file provides summary information about the design and implementation of each table and is very useful for browsing the design of the database. The data-dictionary file, however, does not provide the overview needed to convey the overall design of the database.

A database-design diagram (plate 1) is provided to document selected subject areas in the database and to illustrate relationships between database entities that may not be apparent from examination of the E/R diagrams in this report. This database-design diagram will help the user understand the existing structure and potentially modify the database. This poster-size diagram is 24 by 36 inches and is included as file HRPlate01.pdf on the CD-ROM accompanying this report.

Database Contents

The highway-runoff database has seven general topic areas: descriptions of the data set, highway-runoff monitoring sites, storm events, EMCs, QA/QC data for EMCs, sediment quality, and temporary tables. Most of the tables in the highway-runoff database are used to define storm events and runoff quality. The designs of the data structures for various components of the database are similar to maintain consistency and facilitate understanding and use of the database. The primary criterion for items included in the data structure was that the information would be potentially useful for local, state, regional, or national highway-runoff planning or management efforts. The secondary criterion was the suitability of information that can be stored, searched, and manipulated as plain text or numerical data. The following discussion of each data entity and the associated data structure focuses on the design of the database by topic. Each data structure is documented in an E/R diagram. Tables within each data structure are identified as needed. Detailed table and field definitions, however, are documented in the Microsoft Access database and in the data dictionary on the enclosed CD-ROM.

Data formats such as maps, schematic diagrams, or engineering drawings are not included in the database. The location of some of these elements, however, may be included in comment fields within the database, which could be expanded to include links to electronic files containing these elements. Microsoft Access does support Object Linking and Embedding (OLE) and hyperlink fields so that a database can activate such computer-format files (if all files are copied to maintain the integrity of links and the appropriate software is available on the user's computer). Alternatively, other applications (for example, geographic-information system software) are able to use data in an Access database through an open-database-connectivity (ODBC) driver. Further development of this type of structure would depend on standardization of file formats among state and federal agencies that may use the database information.

Highway-Runoff Data Set

A highway-runoff data set is defined herein as the results of one or more closely related runoff studies that share common methods, materials, and performance measures for the field and laboratory components. Almost all the data sets currently in the database meet this operational definition. The 1990 FHWA runoff-quality model compilation, however, included data from many studies, each with substantially different characteristics (Driscoll and others, 1990c; d).

The highway-runoff data-set data structure provides a means for defining a data set, describing the quality of data in the data set, listing the source documents for the data set, and identifying highway-runoff monitoring sites. This data structure is illustrated in figure 21. The table tblQWHighwayDataSet includes the data-set primary key (QWHighwayDataSet_ID), the data set name, and the period of record. Each data set is associated with one or more citations through the association table (tasHighwayDataSetCitation) to the citation table (tblCitation). The need for such an association documenting the original source of the data was identified in a recent NCHRP study (Venner and others, 2004) as a limitation of the international BMP database and as a research need for development of a highway-runoff quality database. A rudimentary data-quality assessment may be recorded by use of fields in the table tblQWHighwayDataSetQAQC (fig. 21). This basic overview of data quality is not as robust as the reviews done for the NDAMS studies (Dionne and others, 1999; Granato, Dionne, Tana, and King, 2003; Granato 2003), however, the citation table (tblCitation) includes the field Citation_ID, which can be used to identify highway-runoff database citations as entries in the NDAMS database so that data in the highway-runoff database can be integrated with data-quality review results in the NDAMS database. Development of a database with a link between runoff data and the information in the NDAMS database also was identified as a research need in the NCHRP study (Venner and others, 2004). Information about the laboratory or laboratories that analyze samples from a given study may provide critical data-quality information about study results. Information identifying the laboratory or laboratories used for a given study may be documented by use of the table tblQWLab through the association table tasDataSetLab (fig. 21).

Highway-Runoff Monitoring Sites

Highway-runoff monitoring sites are defined herein as sections of road or highway that drain to a specific location from which runoff-flow measurements and water-quality samples are taken. In some studies, one highway-runoff monitoring site may be nested within the drainage area of a separate highway-runoff monitoring site. For example, in the MA 2002 data set, one site is used to monitor runoff draining to a single catch basin and a separate site is used to monitor the inlet of an oil-grit separator (Smith, 2002). The oil-grit separator inlet receives runoff from the first monitoring site and several other (unmonitored) catch basins. Alternatively, it is possible that the same highway-runoff monitoring site could be used in multiple data sets (although site characteristics such as traffic volume, surrounding land use, and pavement materials may change between studies). The database is not currently designed to provide the recursive relationships necessary to make the associations between nested sites or between one site that may be used in multiple studies, but this information may be recorded with the site characteristics in the highway-site narrative memo-field.

The highway-runoff monitoring-site data structure provides a means for defining the characteristics of a monitoring site. This data structure is illustrated in figure 22. The table tblQWHighwayDataSet from figure 21 is included to reinforce the association of each site to its respective data set. Many of the location and site characteristics identified as being important to the identification and interpretation of highway-runoff quality (Driscoll and others, 1990c; d; Thomson and others, 1996; 1997; Granato, Zenone, and Cazenas, 2003) are documented in table tblQWHighwaySite (fig. 22). The user may define the site by highway number, state, county and city, highway mile post, and by decimal latitude and longitude coordinates. Most highway studies identify the site by highway milepost rather than by latitude and longitude coordinates, which hampers the identification of the site location on a regional or national scale. The widespread availability of global-positioning system (GPS) devices will make it easier to include this important data in current and future research studies. Use of detailed, geographically referenced site maps indicating location, highway geometry, drainage structures, and surrounding land use (for example Smith, 2002) would greatly enhance the potential quality of highway-runoff monitoring documentation (Granato, Zenone, and Cazenas, 2003). The table tblQWHighwaySite includes a field to record the accuracy of estimated latitude and longitude coordinates for the highway-runoff monitoring sites.

The domain tables tdxState, tdsPavementType, tdsSectionType, and tdsDrainageSystemType provide standard choices for classifying highway-runoff monitoring site characteristics. Driscoll and others (1990c) indicated that highways designed with curbs or similar structures that route highway runoff along the edge of the pavement to a centralized drainage-collection point had higher concentrations of runoff constituents than highways designed without such structures, so that runoff flows directly off the pavement onto shoulders or median strips. The table tdmYNU (fig. 22) provides a list of standard choices Y (Yes), N (No), and U (Unknown) to classify whether each site may have a curb or berm (a sloped curb) structure to contain runoff at the edge of the pavement. The association table with data tadEventSiteTraffic provides a structure to record storm-by-storm traffic-volume information by use of the one-to-many relationships between tblQWHighwaySite, tblStormEvent, and this table.

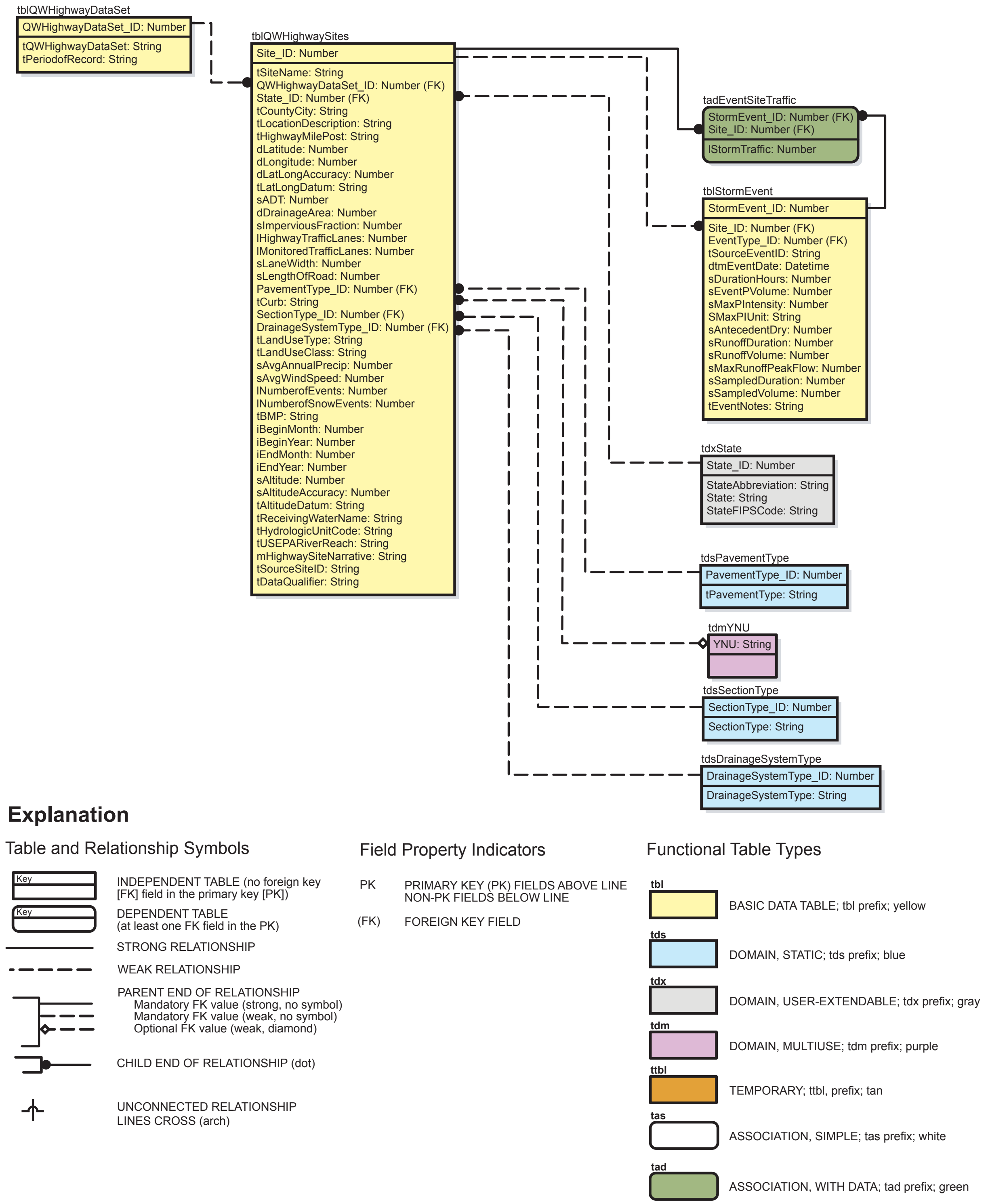


Figure 22. An entity-relationship (E/R) diagram showing a graphical representation of tables, fields, and relationships of the data structure for the highway-runoff monitoring sites.

Highway-Runoff Data

Highway-runoff data are defined herein as the storm-event characteristics and EMC information necessary to estimate concentrations and loads of constituents of interest at a given highway-runoff monitoring site. Storm-event characteristics include information about the antecedent dry period, the date and time of each storm, and precipitation characteristics. The EMC data includes the USEPA parameter information, qualification codes, and the measured or calculated EMC. All constituent concentration data included with the runoff data should represent EMC values because this database was designed to provide planning-level estimates of the population of EMC values for any given site. An EMC is operationally defined as the total storm load of a constituent divided by the total runoff volume. An EMC value may be derived one of three ways (Driscoll and others, 1990c; Strecker and others, 2001; Bent and others, 2003; Breault and Granato, 2003; Lopes and Dionne, 2003). One method is to collect flow-weighted composite samples throughout a storm event and send the resulting composite sample to the laboratory for analysis of an EMC. Another method is to collect time-based samples throughout a storm, composite each subsample by measured flow volume during the constant-time interval and submit the resulting flow-weighted composite for analysis of the EMC. Use of time-based composites can be difficult because of wide variations in precipitation and runoff that may occur between sampling intervals. A third way to estimate an EMC is to submit each subsample for analysis and to multiply each resultant concentration by the flow measured during the sampling interval; the results are then divided by the total runoff volume to estimate the EMC. Theoretically, each of these methods should produce an equivalent EMC value. The comment field in the table tblEMCValues may be used to document the method that is used.

The data structure for the highway-runoff data provides a means for defining the characteristics of the storms and EMC values for each monitoring site. This data structure is illustrated in figure 23. Each storm event in table tblStormEvent and each EMC in table tblEMCValues is identified with a monitoring site by use of the FK Site_ID from table tblQWHighwaySites. The relations from tblQWHighwaySites to these two tables are one-to-many relationships because each site may be associated with zero, one, or more than one storm and zero, one, or more than one EMC measurement. Similarly, the PK StormEvent_ID is a FK in table tblEMCValues, and the one-to-many relationship between these tables allows the user to define zero, one, or more than one constituent analysis for each storm event. The domain table tdsEventType is used to classify the type of precipitation event in table tblStormEvent. The domain table tdsUSEPAParameterCodes defines the parameter code, the parameter name, and measurement units for each EMC value. The database user must be cognizant of these factors when EMC data are input and when extracted for analysis. EMC values are commonly expressed as milligrams or micrograms per liter for whole water and dissolved (filtered) analyses. The domain table tdsAnalysisMethod may be used to identify the laboratory analysis method used to determine the EMC. This information can be useful for assessing the potential effects of current and historic detection and reporting limits and for assessing the potential for systematic bias among different measurement methods. The field tEMCQual is a text field that provides an EMC qualification code allowing the user to identify censored values (Helsel, 2005). The application is implemented on the assumption that all qualified values are left-censored so other qualifications (such as e for estimated) should be entered in the comment field tEMCComment. The domain table tdsCollectionMethod may be used to identify the method for sample collection, which may be used to assess the comparability of EMC values from different data sets.

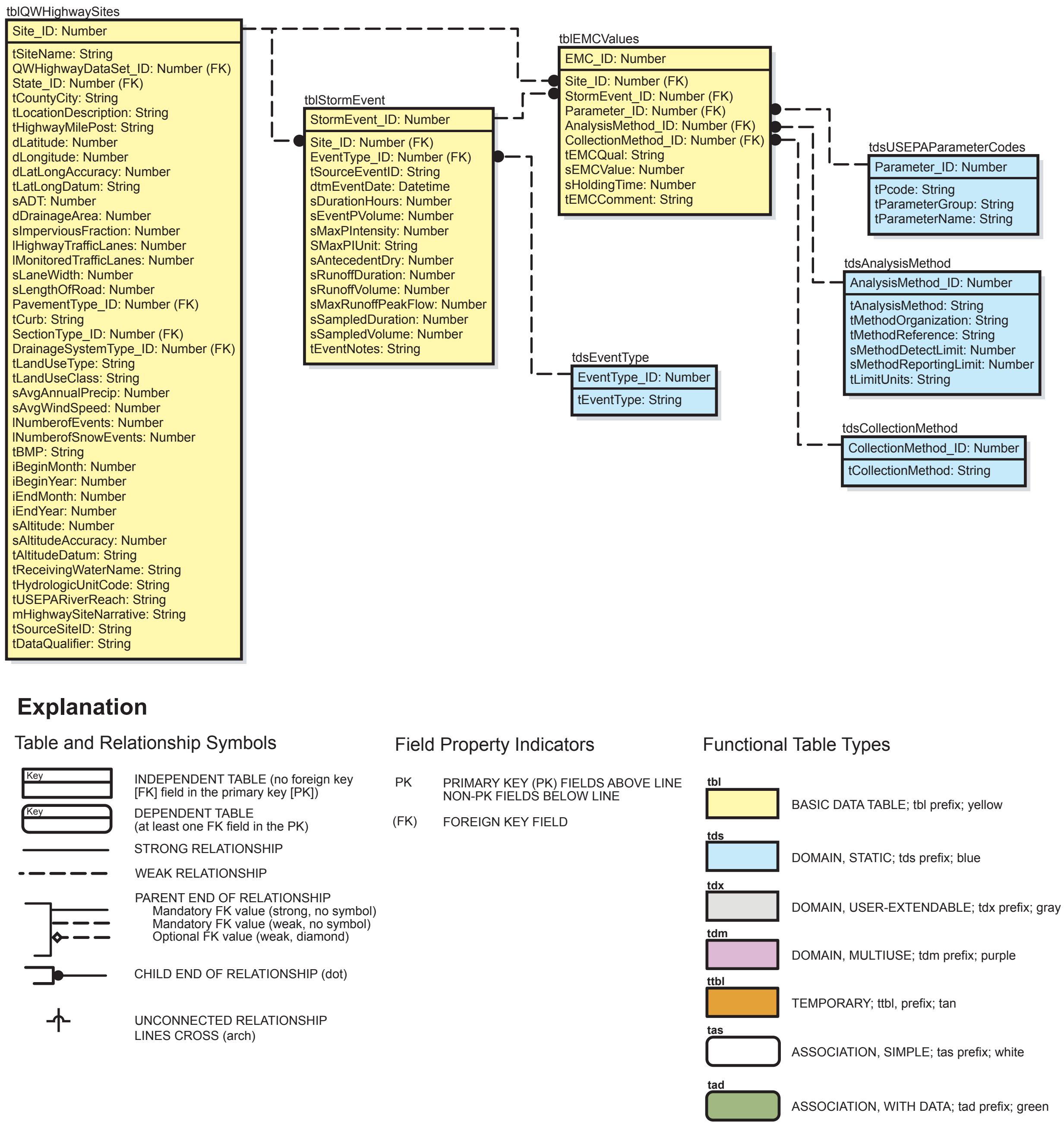


Figure 23. An entity-relationship (E/R) diagram showing a graphical representation of tables, fields, and relationships of the data structure for the highway-runoff monitoring data.

Quality-Assurance and Quality-Control Data for Event Mean Concentrations

QA/QC data are defined herein as the results of chemical analysis necessary to document the quality and potential uncertainty of EMC measurements. As such, the QA/QC data structure (fig. 24) parallels the EMC data structure (fig. 23). There are some differences, however, because QA/QC data commonly are designed to support an entire data set rather than a single storm-event sampling episode. Most of the recent data sets in the database have associated QA/QC data, but entry of these data in the current version of the database was beyond the scope of the current study.

The data structure for the QA/QC data provides a means for defining the characteristics of these data by data set, monitoring site, or EMC value (fig. 24). Each QA/QC sample analysis result is defined as a member of the data set by use of the FK QWHighwayDataSet_ID in a one-to-many relationship from table tblQWHighwayDataSet to table tblQAQCData. A QA/QC measurement in table tblQAQCData may be associated with one or more highway-runoff monitoring sites in table tblQWHighwaySites by use of one-to-many relationships from each table to the association table tasQAQCSite. This structure is necessary because a site may be associated with zero, one, or more than one QA/QC samples. In addition, a QA/QC sample, for example a laboratory blank sample, may be applicable to data from one or more sites.

Similarly, EMC values in table tblEMCValue are related to QA/QC values in table tblQAQCData by use of relationships with the association table tasQAQCEMC. The domain tables tdsUSEPAParameterCodes, tdsAnalysisMethod, and tdsCollectionMethod are used to define both the EMC values and the QA/QC values in the same way. The type of QA/QC sample is further defined in table tblQAQCData by use of a one-to-many relationship from the domain table tdsQAQCSampleType.

Sediment-Quality Data

Sediment-quality data is defined herein as the physical and chemical measurements necessary to document the analysis of the properties and chemistry of street dirt, soil, suspended sediment, and bottom sediment. In a review of reports on potential and ecological effects of highway runoff, Buckler and Granato (2003) indicate that highway runoff is not commonly acutely toxic. However, the review results do indicate that ecological effects such as elevated biological-tissue concentrations, reduced population counts, and reduced species diversity occur in areas where highway sediments accumulate. Chemical analysis of sediment samples is an efficient way to determine the sources, transport, and fate of trace metals, trace organic compounds, and, potentially, other constituents transported in highway runoff (Breault and Granato, 2003; Lopes and Dionne, 2003). Finally, chemical analysis of sediment samples may provide the information necessary to estimate concentrations of these constituents reported as being below detection limits in whole-water EMC samples. Some studies in the database have associated sediment-quality data, but entry of these data in the current version of the database was beyond the scope of the current study.

The sediment-quality data structure (fig. 25) is designed to be flexible enough to record a wide variety of sediment analyses that are commonly associated with urban and highway stormwater studies. A sediment sample may be collected from one or more sampling locations in or near a highway monitoring site. These sample locations may include the pavement, catch basins, drainage pipes, BMP structures, receiving waters, and other paved and unpaved areas inside and outside a highway right-of-way. Therefore, table tblQWHighwaySites, which is used to identify the highway characteristics, has a one-to-many relationship with tblSedimentSamplingSite, which is used to document the location and description of one or more associated sediment-sampling sites. The domain table tdsSedSampleType is used to classify such sites to allow comparison of data from similar locations at different highway sites. The table tblSedimentSample is used to define the bulk-sample properties. The domain tables tdsSedSampleMatrix, tdsSedSampleType, tdsSedSampleMethod are used to classify the sample collection matrix, type, and collection method, respectively.

The table tblSedimentSample is associated with the table tblStormEvent through the association table tasSedimentStorm (fig. 25). This is because each sediment-quality sample may be associated with one or more storm events. For example, a sediment-quality sample collected from a BMP representing sediment accumulations that occur over one or more storms. Conversely, one or more sediment-quality samples may be associated with each storm event. For example, a suspended-sediment sample could be collected from the pavement, a catch basin, and the outfall to a receiving-water body from a given storm for the purpose of sediment-quality analysis. It also is possible that a sediment-quality sample may not be associated with any particular storm event. For example, a transportation agency may collect sediment-quality samples from catch basins, structural BMPs and (or) the outfall to receiving waters at sites with and without storm-runoff monitoring data. Sediment-quality data may be used as a reconnaissance tool because these data are, in comparison to a runoff-monitoring study, relatively easy and inexpensive

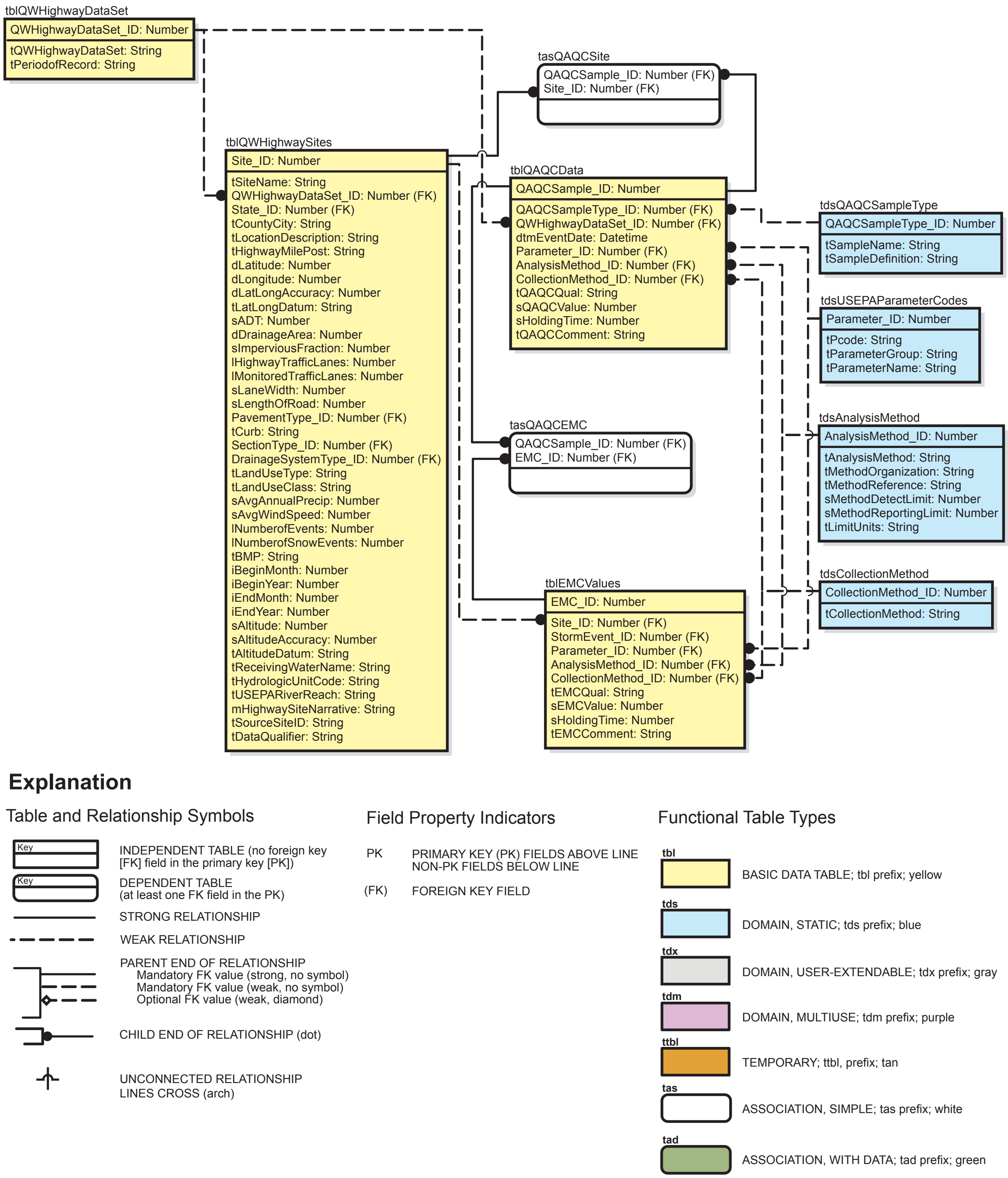


Figure 24. An entity-relationship (E/R) diagram showing a graphical representation of tables, fields, and relationships of the data structure for the quality-assurance and quality-control data for event mean concentrations.

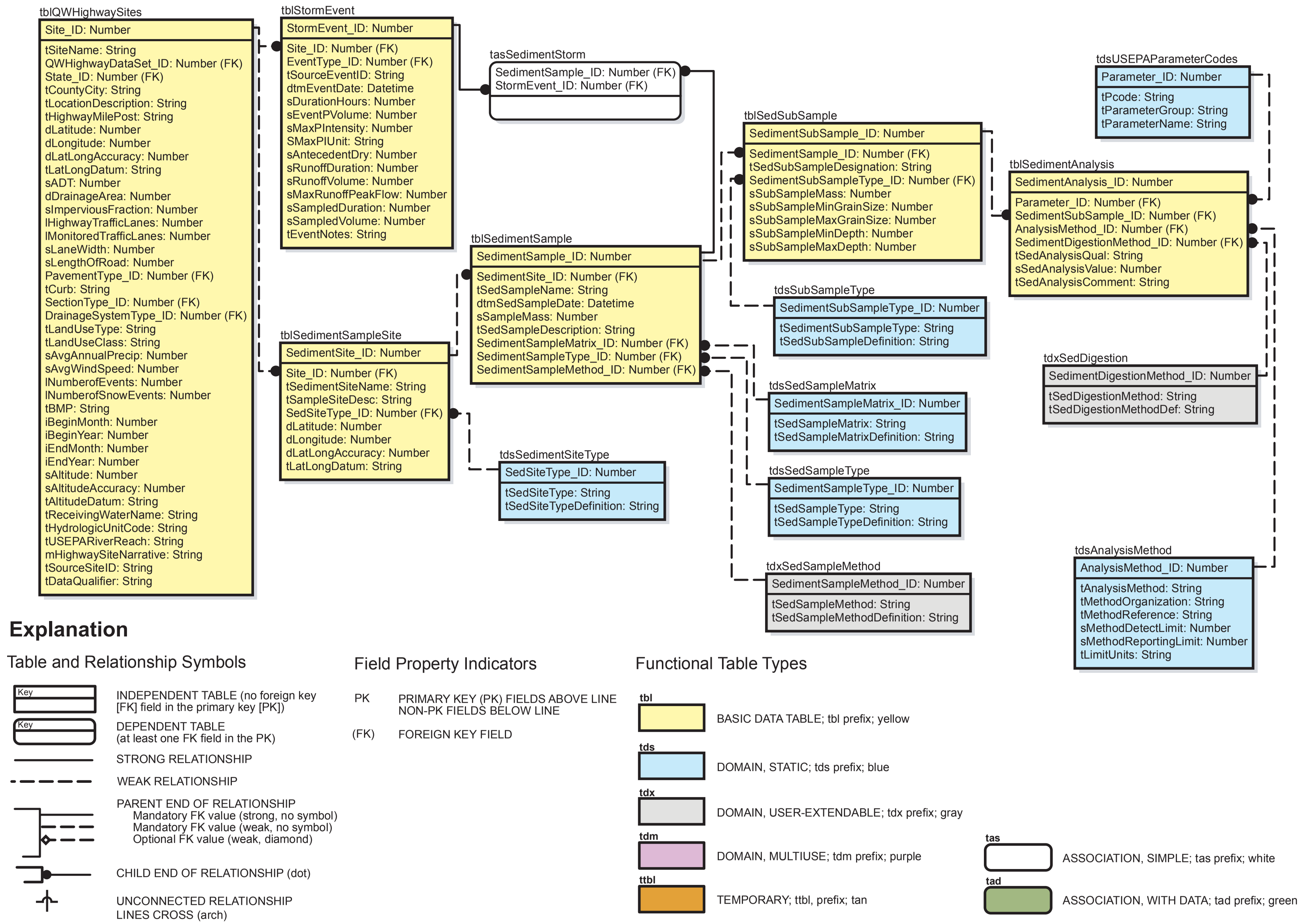


Figure 25. An entity-relationship (E/R) diagram showing a graphical representation of tables, fields, and relationships of the data structure for the sediment-quality data.

to collect and analyze. Sediment-quality data could be used for selection of multiple highway-runoff monitoring sites that have range of expected runoff quality or be used as an indicator of how representative data from a monitored site may be for application to other unmonitored sites.

Sediment samples are commonly split into subsamples by grain size and (or) by sediment-core depth prior to analysis. Therefore, the table tblSedimentSample has a one-to-many relationship with tblSedSubSample (fig. 25) that may be used to record the properties of a single bulk sample, or a number of subsamples. The domain table tdsSubSampleType is used to define the sediment subsample as an environmental sample or a QA/QC sample.

Each subsample may be analyzed for one or more sediment analytes. Therefore, the table tblSedSubsample has a one-to-many relationship with the table tblSedimentAnalysis (fig. 25). Each chemical analysis can be defined by parameter (tdsUSEPAParameterCodes), sediment digestion method (tdxSedDigestion), and laboratory analysis method (tdsAnalysisMethod). The domain table tdsUSEPAParameterCodes defines the parameter code, the parameter name, and measurement units. The database user must be cognizant of these factors when the date are input and when extracted for analysis. For example, if the objective is to estimate EMC values for a constituent, the concentration of the constituent of interest may be estimated on the assumption that the sediment fraction accounts for a majority of the constituent of interest (Breault and Granato, 2003; Lopes and Dionne, 2003). The suspended-sediment concentrations are expressed as milligrams or micrograms per liter in the water column and so are implicitly based on the suspended-sediment concentration from a whole-water sample. The concentrations of an analyte in the bed-sediment, however, are commonly expressed as milligrams or micrograms per kilogram of the sediment. Therefore, the analyte concentration in the sediment must explicitly be used in conjunction with the suspended-sediment concentration in the water to estimate the associated concentration of the analyte in a whole-water sample. Furthermore, if grain-size-based subsamples are analyzed, the whole-water concentration of the analyte must be estimated on the basis of the mass fraction of each grain-size class in the runoff sample. The field tSedAnalysisQual is a text field that provides a qualification code allowing the user to identify censored values (Helsel, 2005). The application for EMCs is implemented on the assumption that all qualified values are left-censored; to keep sediment-quality data consistent with the EMC data, other qualifications (such as e for estimated) should be entered in the comment field tSedAnalysisComment.

Temporary Tables

The database currently contains three temporary tables that are required by the database application for the temporary storage and manipulation of data. The tables are actually permanent within the database, but the contents of these tables are temporary because the contents are deleted when the database application loads and when they are about to be reloaded for use. The tables ttblMyTempDataSets and ttblMyTempDataSites temporary store the user-selected identification number(s) for the data set(s) and data-monitoring site(s), respectively. Although these two identification numbers are FKs from the parent data set (tblQWHighwayDataSet) and site (tblQWHighwaySites) tables, and in effect have a one-to-many relationship from the parent table to the child table, these tables are not shown in the database diagram because of the role they play in data manipulation rather than data storage. The temporary table ttblMyROS, also is not displayed in the database diagrams because of the temporary role it plays in data manipulation. Table ttblMyROS does not have any relationships or key fields because it holds a temporary array of sorted values, including the reported EMC values, EMC qualification codes, reported or (if the EMC is below a detection limit) estimated EMC values determined by use of the ROS method described in appendix 1 (Helsel and Cohn, 1988; Driscoll and others, 1990c; Helsel, 2005), the estimated plotting-position values, and the estimated normal scores for all values in a user-selected data set. All these values are not automatically output for the user in the database application interface, but the user may open the table and export these values for further analysis while the statistics form is still open. The temporary tables are shown on plate 1 without their relationships and are documented in the data dictionary file (HRDDv01.pdf) on the CD-ROM accompanying this report.

Operational Issues and Procedures

The highway-runoff database is designed so that users may easily augment the data. The user-interface application automates many common tasks for data manipulation and retrieval. It is, however, necessary to understand the basic data model, details of the data entities, and the operational aspects of working with data in a relational database to make full use of the database content, to add data, and to expand the database. The main operational issues that need to be considered during use of the highway-runoff database include

(1) the need to standardize and control key assignments in the database structure, (2) the need to follow a predetermined table-loading order, (3) methods for customizing and extending the data structure, and (4) simplification of multi-table structures for handling and presenting data.

Key Assignments and Control

The highway-runoff database uses AutoNumber fields for the critical PK values. AutoNumber fields are long integers that automatically are incremented by Microsoft Access to ensure that each new record has a unique identifying key. The highway-runoff database does not use information-rich keys (fields that have code names that may apparently provide unique values) because such keys could be reassigned in practice and would therefore corrupt associated data in other tables. Use of autonumber keys is recommended as a standard relational-database design practice, but this design convention does have a potential liability. Most databases are designed for a specific user and are controlled and maintained by a single organization in a central location (for example, the USGS NWIS). Alternatively, some databases are designed to convey information in a distributed format to many users. These databases are commonly provided as a tool to document results of a data-collection effort and are fully designed and populated when the data is distributed. Therefore, key assignments are not a factor in the distribution design. Ultimately, the FHWA and state transportation agencies may choose a centralized model, a distributed model, an integrated model, or some combination thereof for database implementation.

In a centralized model, the FHWA or some designated organization would host the only official version of the highway-runoff database. The International BMP database (U.S. Environmental Protection Agency, 1999) was designed and implemented by use of such a centralized model. Use of a centralized database, however, may not meet individual data needs beyond the information provided by a standardized nationwide interface. The centralized model also requires a database administrator to check and enter data, maintain the database, and provide information to users. Key assignments and controls would not be a critical factor because each database entry would automatically generate a unique key-field value. The centralized model would allow for analysis of data from individual researchers and for integration of information from different highway-runoff studies without duplication of effort in building data queries. The database administrator would implement standard methods to check, enter, query, and report data.

In a distributed model, each researcher would maintain and control their own version of the highway-runoff database. In this model, key assignments and controls would not be a critical factor because each copy of the database would contain unique and independent information. The power of the relational database could be used to examine information for each researcher through time, but state and federal agencies would not be able to easily integrate and compare information among runoff studies because results from each study would be in a different database file with different key fields. In the distributed model, some data may be lost through time because there is no central archive. The distributed model also could create situations where duplication of effort by different agencies would be necessary to examine the data. Each database owner would be responsible for the quality and consistency of data and information in the distributed model.

In an integrated model, each agency would maintain and control their version of the highway-runoff database, but the different copies of the database would be integrated by appending information and data from each study into a central copy of the database. In this model, researchers would enter data on their own computer systems and send a copy to the FHWA for integration into the central version of the database. Key assignments and controls would be a critical implementation factor for this model. Each copy of the database would be owned and operated by the researchers and would contain unique and independent information. The process of integrating individual copies of the database from each data supplier into the central version, which is owned and operated by the FHWA, could corrupt the database if autonumbers used in the individual copies are not handled properly. This, however, may be addressed by assigning ranges of autonumber key field values to each study upon its inception or by systematically renumbering key fields when data are loaded into the central database. Key-field assignments would be preferable for maintaining consistency among database versions. This method is feasible because the long-integer data type used for keys within Access may range up to a value of 2,147,483,647 (Roman, 1997), which is almost 200,000 times the number of EMCs currently in the database. Therefore, large blocks of autonumber values could be assigned to each transportation agency for future use. The Microsoft Access help file includes detailed instructions for changing the starting value of an autonumber key field.

An integrated model would allow the FHWA, state transportation agencies, and individual highway research organizations to use and compare a common set of information and data. An integrated model also would allow individual data suppliers to design and implement custom queries and reports for their own needs. In an integrated model, the FHWA, state transportation agencies, and individual data suppliers would share responsibilities for quality and consistency of data and information in the individual and the centralized versions of the highway-runoff database. An integrated highway-runoff model also would require some level of database administration to coordinate loading and distributing new database versions.

Table-Loading Order

The design and implementation of the highway-runoff database make it necessary to follow a predetermined loading order. The loading-order information is necessary for manual entry or automated entry of data and for design of a user interface. The loading order is determined by design factors, such as the division of information among tables in each data structure, by implementation factors such as restrictions caused by data-protection settings, and by the use of foreign keys in association tables. The domain tables (tds, tdr, and tdx) are populated with standard choices for database use. The loading order of basic data tables (tbl) depends on the presence of foreign keys. Almost all information in the database is ultimately associated with a highway-runoff data set. Once the data set is established and defined, the user must define monitoring sites. One or more monitored storms may be defined for each data-monitoring site. EMC values are defined by site and storm. Association tables (tas and tad) are dependent on foreign keys from basic data tables and domain tables, and therefore are commonly the last tables to be populated within each data structure. The individual data-structure-design diagrams (figs. 21-25) map the design of the database and may be used to determine the proper table-loading order. Relationships in the highway-runoff database are almost exclusively one-to-many relationships. Parent tables must be loaded first so that the parent information is available for selection in the child table. A database user can determine the table-loading order by examining the appropriate design diagram(s) for each table and data structure.

The database fields are designed and implemented with MS Access "combo boxes," which provide a pull-down list of choices to populate information in a receiving table. These combo boxes provide a pull-down list of choices for each FK field. If the desired choices are not available within a given combo box, the user may not have followed the optimal table-loading order. Other tables must be populated for the desired choices to appear in the combo box. If the appropriate selection is not available in the combo box, the user may consult the appropriate design diagram to follow the relations back to the parent table.

Customizing and Extending the Data Structure

The highway-runoff database was designed as a preliminary structure to provide a basis for collecting and compiling data from runoff studies and as a preprocessor for SELDM. The database was developed to provide an initial design that could be customized, extended, or even truncated as the FHWA, state transportation agencies, regulators, and the research community come to consensus on the type and format of data necessary to meet information needs. The database may be extended by adding fields or by adding tables. To maintain data integrity, these extensions should be done by use of existing conventions that apply to normalization, keys and relationships, domain-table usage, and naming rules. If an individual organization wishes to add auxiliary fields (for example, to record more detailed engineering information about highway sites), a new table that contains the new data fields in a one-to-one relationship with the parent table could be added. In this way, users can enter the desired information without compromising the ability to integrate data from different versions of distributed copies of the database. If the FHWA determines that the new fields are useful to the majority of users and that other database users would use the new information, new tables could be consolidated with existing tables in a future revision. Custom tables and fields should be distinguished with some form of unique identifier. For example, users may designate custom accessory table and field names by a prefix with the letter "z" or other unique indicator (for example, the two-letter postal code identifying the state of the originating DOT) to distinguish them from the standard database objects. If tables, fields, are removed from the highway-runoff database, users should take great care not to corrupt the database by loss of keys and relationships. Similarly removing or renaming tables, fields, or relationships will probably corrupt the database application.

Simplification of Multi-Table Structures

The highway-runoff database was designed with many tables and relationships to maximize data integrity through normalization of information by data set, monitoring site, and storm event. The number of tables and the apparent complexity of relationships between the tables could have the potential to confuse those not familiar with the design. For example, relationships between QA/QC data and the data set, monitoring site, and storm is accomplished through use of one data table with a FK field to the data set, two association tables (to the site and to the storm EMC for a constituent), and four domain tables (fig. 24). Although this may appear to be a complex design, upon further consideration it becomes apparent that a QA/QC sample may be analyzed for the entire study (for example, a sample used to test laboratory variability), for the monitoring site (for example, a sample used to test for contamination from sampling equipment), or be associated with an individual EMC (for example, a replicate sample to test sampling and measurement variability). Thus, the complex multi-table design is necessary to allow for likely uses of the data.

Table 1. Example queries available in the highway-runoff database.

[EMC, Event Mean Concentration; FHWA, Federal Highway Administration; KTRLine, Kendall-Theil Robust Line; PCODE, U.S. Environmental Protection Agency Parameter Code; R_v, Runoff coefficient; SELDM, Stochastic Empirical Loading and Dilution Model; TSS, Total Suspended Solids; VSS, Volatile Suspended Solids]

Query name	Group	Purpose
qryDataSetCitation	Data Set	Citations for data in the database by data set
qryDataSetCountSites	Data Set	Count number of monitoring sites grouped by data set
qryDataSetCountStorms	Data Set	Count number of storm events grouped by data set
qryDataSetCountSummary	Data Set	Summary of data-set count queries. Example of compilation of information by combining queries.
qryDataSetTimeLine	Data Set	Sampling time-line by data set
qrySiteStorms	Site	Count of storms with begin date and end date by site
qrySiteSummary	Site	Site information in format for tblQWHighwaySite in SELDM
qryCountEMCbyGroup	Water Quality	Count number of EMC measurements by parameter group
qryCountEMCByParam	Water Quality	Count number of EMC measurements grouped by parameter and sorted by group and PCODE
qryCountEMCbyPCODEbyDtaSet	Water Quality	Count number of EMC measurements grouped by data set and sorted by PCODE
qryCountParambyEMC	Water Quality	Count EMC measurements grouped by parameter and sorted by group and PCODE
qryDataSetCountEMC	Water Quality	Count number of EMC measurements grouped by data set
qryCountEMCPairedWithTSS	Water Quality	Count of EMC measurements paired with TSS. Example of a self join on a table for recursive relation
qryCountPairedEMCValues	Water Quality	Count of paired EMC measurements (TSS & VSS) example of a self join on a table for recursive relation
qryCalculateRv	Runoff Coefficient	Calculate runoff coefficient by storm and by site
qryCalculateRv4KTLine	Runoff Coefficient	Runoff coefficient information in KTRLine format
qryGetRv01	Runoff Coefficient	Runoff coefficient with estimated area from the 1990 FHWA R_v regression equation
qryRvStatsBySite	Runoff Coefficient	Runoff coefficient statistics by site

Data users commonly want a simplified one-table custom view of the data and different users commonly want different views of the data. For example, a contracting official may want a count of each QA/QC sample type to see if requirements have been met, whereas a scientist may want one query with individual values for each constituent and QA/QC sample type to determine the confidence interval of reported values. Similarly, a QA/QC specialist may want individual concentration values coupled with dates to detect performance trends through time. In short, it is difficult or impossible to anticipate all potential views for the data.

In practice, database users can build queries that combine information from multi-table structures into a composite view of database contents. The resulting query can be manipulated and used like a single table in the database without duplication of information or loss of normalization. For example, the query (qryGetRv01) aggregates information from the site table (tblQWHighwaySites) and the storm-event table (tblStormEvent) and converts measurements to a common set of units (feet of precipitation, square feet of drainage area, and cubic feet of runoff) to calculate a runoff coefficient for each storm at each site (table 1). This query also demonstrates the principle that calculated fields, such as runoff coefficients or storm loads, can be generated by a query rather than by duplicating data in the database tables. The highway-runoff database includes this query and a number of other example queries (table 1) that provide useful information, examples of data-aggregation techniques, and demonstrations of methods for simplifying multi-table structures by combining data from related tables.

Summary

Development of an up-to-date database of highway-runoff data has been identified as a long-standing, high-priority need for environmental research. Knowledge of the quality and quantity of highway runoff and associated sediments is important for decision makers, planners, and highway engineers to assess and mitigate possible adverse effects of highway runoff on the Nation's receiving waters. Data and information about precipitation, and the quality and quantity of highway runoff from sites with different highway design characteristics, traffic volumes, and surrounding land uses may help define variations in runoff quality from site to site. Data and information from different parts of the country may be used to characterize highway-runoff quality as a function of regional variations in fuel formulations, emission standards, construction and maintenance practices, and variations in soil geochemistry. Highway-runoff data also are necessary to assess the effectiveness of potential Best Management Practices (BMPs). Finally, such data are necessary to formulate planning-level estimates of runoff quality for existing or planned highway-runoff sites with insufficient monitoring data.

The highway-runoff database (HRDB) application was developed by the U.S. Geological Survey, in cooperation with the Federal Highway Administration (FHWA), to serve as a data warehouse to document data and information from available highway-runoff monitoring studies. The HRDB application also is designed to be a preprocessor to produce statistics for runoff coefficients and event mean concentration (EMC) measurements for use with the Stochastic Empirical Loading and Dilution Model (SELDM) that is designed to update and improve the 1990 FHWA runoff-quality model. To meet data needs, the highway-runoff database was populated with 39,713 EMC measurements (including 116 different water-quality constituents), from 2,650 storm events, monitored at 103 highway-runoff monitoring sites in the conterminous United States, as documented in 7 highway-runoff quality data sets. This HRDB application provides the information necessary to estimate water-quality statistics on the basis of highway-site characteristics, to define surrogate parameter relations (such as an equation to predict suspended-sediment concentrations from total suspended solids), and to estimate runoff coefficients on the basis of impervious fraction of a given site. This database application provides the information and data necessary to explore relations between measured runoff quality and various explanatory variables.

Step-by-step use of the graphical-user interface for the HRDB application is described in this report. The user has five options to use the data in the database through this interface. The user may select and export: (1) water-quality data in tab-delimited format; (2) water-quality data in a format suitable for analysis with censored-data programs; (3) paired water-quality data in a format suitable for regression analysis; (4) water-quality statistics by use of the robust regression on order statistics (ROS) method; and (5) precipitation, runoff flow, and runoff-coefficient data in tab-delimited format. In each case, the user may select data from different sites and different data sets based on highway-runoff monitoring site characteristics. These five options are provided so that the user may select, from all available data, a custom data set and do the analysis necessary to estimate the properties of runoff quality and flow that are representative of a site of interest. In this way, the database may be used to define highway runoff at monitored sites, and estimate highway-runoff properties at unmonitored sites.

The database file contains 37 tables, including 6 association tables, 12 data tables, 16 domain tables, and 3 temporary-data tables. This report, the database dictionary, the database design diagram, and the database file on the CD-ROM accompanying this report document the design and contents of the database. Information about the design and implementation of the application and underlying database are provided to facilitate future use and modification of the HRDB application. The program code, written in Microsoft Visual Basic for applications, is documented within the Microsoft Access database file on the CD-ROM accompanying this report. Some basic information about database design and implementation in Microsoft Access is provided. The report, however, was written with the assumption that the intended audience for the implementation and design portion of this document has a working knowledge of Microsoft Access and some background in the design or use of relational databases. Information and training on the use of Microsoft Access is widely available and can be located on the Internet.

The highway-runoff database is designed so that users may easily augment the data. The user-interface application automates many common tasks for data manipulation and retrieval. It is, however, necessary to understand the basic data model, details of the data entities, and the operational aspects of working with data in a relational database to make full use of the database design. The main operational issues which need to be considered during use of the database include (1) the need to standardize and control key assignments within the database structure, (2) the need to follow a predetermined table-loading order, (3) methods for customizing and extending the data structure, and (4) simplification of multi-table structures for handling and presenting data. These issues are described as operational issues and procedures for current and future use of the highway-runoff database application.

Acknowledgments

The information and data are available in this highway-runoff database because of the contributions of a number of highway-runoff researchers. Eric Strecker and Eugene Driscoll, authors of the 1990 FHWA runoff-quality model study (Driscoll and others, 1990a; b; c; d), provided data and information about their compilation methods including information about detection-limit data in the 1990 working-data spreadsheet files. Michael Trouchon and Tod Granicher provided information and data for the California highway-runoff data set. Kirk Smith, Molly Lamrouex, Michael Barrett, Richard Tveten, and Robert Waschbusch provided information and data for the Massachusetts, Michigan, Texas, Washington State, and Wisconsin highway-runoff data sets, respectively. Kirk Smith and Robert Breault, USGS, provided information for design of the sediment-quality data structure. Dennis Helsel, USGS, provided information about the statistical treatment of detection-limit data and the FORTRAN source code for the USGS MDL software. Dylan Begin, a USGS Volunteer for Science, translated the site descriptions from the 1990 FHWA runoff-quality model data report and original documents into estimates of latitude and longitude for these sites by use of the USGS National Map (U.S. Geological Survey, 2005). Leslie DeSimone, USGS water-quality specialist, and David A. Graves, New York Department of Transportation water-quality specialist, provided detailed technical reviews of the highway-runoff database and the database documentation.

References Cited

American Public Health Association, American Water Works Association, and Water Pollution Control Federation, 1995, Standard methods for the examination of water and wastewater (19th ed.): Washington, D.C., American Public Health Association, variously paged.

American Society for Testing and Materials, 2000, Standard test method for determining sediment concentration in water samples: ASTM Designation D 3977-97, p. 395–400.

Athayde, D.N., Shelly, P.E., Driscoll, E.D., Gaboury, D., and Boyd, G., 1983, Results of the nationwide urban runoff program, v. 1—Final report: U.S. Environmental Protection Agency, Water Planning Division Report WH-554, 186 p.

Bank, F.G., 1993, Water quality research needs in transportation, *in* Session #165, Heightened awareness of highway water quality management: Washington, D.C., Transportation Research Board Annual Meeting, January 10-14, 7 p. (also at URL: http://www.fhwa.dot.gov/environment/wqrsneed.htm)

Bank, F.G., Kerri, K.D., Young, G.K., Stein, Stuart, 1996, National evaluation of water quality issues for highway planning: Accessed on the internet on July 15, 2005 at URL http://www.fhwa.dot.gov/environment/wqnatevl.htm

Barrett, M.E., Malina, J.F., Jr., Charbeneau, R.J., and Ward, G.H., 1995, Characterization of highway runoff in the Austin, Texas area: Austin, TX, University of Texas at Austin, Bureau of Engineering Research, Center for Research in Water Resources Report CRWR 263, 35 p.

Barrett, M.E., Malina, J.F., Jr., Charbeneau, R.J., and Ward, G.H., 1996, Characterization of highway runoff in the Austin, Texas area: Texas Department of Transportation Report FHWA/TX-96/1943-4, 38 p.

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, v. 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.

Blom, G., 1958, Statistical estimates and transformed beta variables: New York, John Wiley, p. 68–75, 143–146.

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, v. 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. 165–234.

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, v. 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.

California Department of Transportation, Office of Environmental Engineering, 2000, Guidance manual—Stormwater monitoring protocols (2nd ed.): Sacramento, CA, California Department of Transportation Report CTSW-RT-00-005, 303 p.

California Department of Transportation, Office of Environmental Engineering, 2002, CALTRANS statewide stormwater runoff characterization study—Monitoring season 2001-2002 final report: Sacramento, CA, California Department of Transportation, Report CTSW-RT-02-022, 583 p.

California Department of Transportation, Office of Environmental Engineering, 2003a, Storm water monitoring and research program preliminary report of discharge characterization studies: Sacramento, CA, California Department of Transportation Report CTSW-RT-03-023, 42 p.

California Department of Transportation, Office of Environmental Engineering, 2003b, CALTRANS Tahoe highway runoff characterization and sand trap effectiveness studies 2000-03 monitoring report: Sacramento, CA, California Department of Transportation Report CTSW-RT-03-054.36.02, 895 p.

California Department of Transportation, Office of Environmental Engineering, 2003c, Storm water monitoring and data management discharge characterization study report: Sacramento, CA, California Department of Transportation Report CTSW-RT-03-065.51.42, 93 p.

California Department of Transportation, Office of Environmental Engineering, 2003d, Storm water monitoring and data management 2002-2003 annual data summary report: Sacramento, CA, California Department of Transportation Report CTSW-RT-03-069.51.42, 1859 p.

California Department of Transportation, Office of Environmental Engineering, 2004, Storm water monitoring and BMP development status report: Sacramento, CA, California Department of Transportation Report CTSW-RT-03-069.51.42, 111 p.

CH2MHill Inc., 1998, Highway stormwater runoff study: Lansing, MI, Michigan Department of Transportation Report MDOT-MS4-91946-7, variously paged.

Chow, V.T., 1954, The log-probability law and its engineering applications: Proceedings of the American Society of Civil Engineers, v. 80, separate 536, 25 p.

Church, P.E., Granato, G.E., and Owens, D.W., 2003, Basic requirements for collecting, documenting, and reporting precipitation and stormwater-flow measurements, *in* Granato, G.E., Zenone, C., and Cazenas, P.A. (eds.), National highway runoff water-quality data and methodology synthesis, v. 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. 47–79.

Cohn, T. A., 1988, Adjusted maximum likelihood estimation of the moments of lognormal populations from type I censored samples: U. S. Geological Survey Open-File Report 88-350, 34 p.

Cunnane, C., 1978, Unbiased plotting positions—A review: Journal of Hydrology, v. 37, p. 205–222.

Dionne, S.G., Granato, G.E., and Tana, C.K, 1999, Method for examination and documentation of basic information and metadata from published reports relevant to the study of stormwater runoff quality: U.S. Geological Survey Open-File Report 99-254, 156 p.

Driscoll, E.D., Palhegyi, G.E., Strecker, E.W., and Shelley, P.E., 1989, Analysis of storm event characteristics for selected rainfall gauges throughout the United States: Washington, D.C., U.S. Environmental Protection Agency, 43 p.

Driscoll, E.D., Shelley, P.E., and Strecker, E.W., 1990a, Pollutant loadings and impacts from highway stormwater runoff, v. I—Design procedure: Washington, D.C., 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, v. II—User's guide for interactive computer implementation of design procedure: Washington, D.C., 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, v. III—Analytical investigation and research report: Washington, D.C., 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, v. IV—Research report data appendix: Washington, D.C., Federal Highway Administration Final Report FHWA-RD-88-009, 143 p.

Fleming, C.C. and von Halle, Barbara, 1989, Handbook of relational database design: Reading, Mass., Addison-Wesley Publishing Company, 605 p.

Glysson, G.D., Gray, J.R., and Conge, L.M., 2000, Adjustment of total suspended solids data for use in sediment studies, *in* American Society of Civil Engineers 2000 Joint Conference on Water Resources Engineering and Water-Resources Planning and Management, Minneapolis, MN, July 31–August 2, 2000, Proceedings: Reston, VA, 10 p.

Granato, G.E., 2003, National highway runoff water-quality data and methodology synthesis, v. III—Availability and documentation of published information for use in regional or national highway-runoff quality data synthesis: Washington, D.C., U.S. Department of Transportation, Federal Highway Administration, FHWA-EP-03-056, 71 p.

Granato, G.E., 2004, Rhode Island water supply system management plan database (WSSMP–Version 1.0): U.S. Geological Survey Open-File Report 2004-1231, 77 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. with CD-ROM.

Granato, G.E., Bank, F.G., and Cazenas, P.A., 1998, Data-quality objectives and criteria for basic information, acceptable uncertainty, and quality-assurance and quality-control documentation: U.S. Geological Survey Open-File Report 98-394, 17 p.

Granato, G.E., Dionne, S.G., Tana, C.T., and King, T.L., 2003, National highway runoff water-quality data and methodology synthesis, v. II—Project documentation: Washington, D.C., U.S. Department of Transportation, Federal Highway Administration, FHWA-EP-03-055, 22 p. with CD-ROM.

Granato, G.E., and Tessler, Steven, 2001, Data model and relational database design for highway runoff water-quality metadata: U.S. Geological Survey Open-File Report 00-480, 15 p.

Granato, G.E., Zenone, C., and Cazenas, P.A. (eds.), 2003, National highway runoff water-quality data and methodology synthesis, v. 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, 479 p.

Gray, J.R., Glysson, G.D., Turcios, L.M., and Schwarz, G.E., 2000, Comparability of suspended-sediment concentration and total suspended solids data: U.S. Geological Survey Water-Resources Investigations Report 00-4191, 14 p.

Gringorten, I. I., 1963, A plotting rule for extreme probability paper: Journal of Geophysical Research, v. 68, p. 813–814.

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

Hazen, A., 1914, Storage to be provided in the impounding reservoirs for municipal water supply: Transactions of the American Society of Civil Engineers, v. 77, p. 1547–1550.

Helsel, D.R., 2004, Practical stats—Practical statistics for the sciences: Accessed on the internet on June 21 2005 at URL: http://www.practicalstats.com

Helsel, D.R., 2005, Nondetects and data analysis: New York, Wiley, 250 p.

Helsel, D.R., and Cohn, T.A., 1988, Estimation of descriptive statistics for multiply censored water quality data: Water Resources Research, v. 24, no. 12, p. 1997–2004.

Helsel, D.R., Cohn, T.A., and Harrington, John, 1988, Program to estimate summary statistics for data with multiple detection limits for Microsoft Windows (MDLWIN): Accessed on the internet on June 21 2005 at URL: http://www.practicalstats.com

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.

Hernandez, M.J., 1997, Database design for mere mortals—A hands-on guide to relational database design: Reading, MA, Addison-Wesley Publishing Company, 480 p.

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.

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, v. 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.

National Institute of Standards and Technology, 1993, Standard for integration definition for information modeling (IDEF1X): Federal Information Processing Standards Publication 184, 155 p.

Roman, Steven, 1997, Access database design and programming: Sebastopol, CA, O’Reilly and Associates, 251 p.

Schueler, T.R., 1987, Controlling urban runoff—A practical manual for planning and designing urban BMP’s: Washington, D.C., Metropolitan Washington Council of Governments, Department of Environmental Programs, 275 p.

Shumway, R.H., Azari, R.S., and Kayhanian, M., 2002, Statistical approaches to estimating mean water quality concentrations with detection limits: Environmental Science and Technology, v. 36, no. 15 p. 3345–3353.

Smith, K.P., 2002, Effectiveness of three best management practices for highway-runoff quality along the Southeast Expressway, Boston, Massachusetts: U.S. Geological Survey Water-Resources Investigations Report 02-4059, 62 p., + 1 CD-ROM.

Strecker, Eric; Mayo, Lynn; Quigley, Marcus; Howell, James, 2001, Guidance manual for monitoring highway runoff water quality: Washington, D.C., U.S. Department of Transportation, Federal Highway Administration, FHWA-EP-01-022, 206 p.

Strecker, Eric; Urbonas, Ben; Quigley, Marcus; Howell, James; and Hesse, Todd, 2002, Urban stormwater BMP performance monitoring—A guidance manual for meeting the national stormwater BMP database requirements: Washington, D.C., U.S. Environmental Protection Agency Report EPA-821-B-02-001, 248 p.

Taylor Associates, Inc., 2002a, SR 167 Ecology embankment water quality monitoring project final report: Olympia, WA, Washington State Department of Transportation, 28 p.

Taylor Associates, Inc., 2002b, SR 405 Vortechs water quality monitoring project final report: Olympia, WA, Washington State Department of Transportation, 267 p.

Technology Acceptance and Reciprocity Partnership, 2001, Stormwater best management practice demonstration tier ii protocol for interstate reciprocity—Endorsed by California, Massachusetts, New Jersey, Pennsylvania, and Virginia: Technology Acceptance and Reciprocity Partnership, variously paged.

Tetra Tech, Inc., 2002, National pollutant discharge elimination system, Washington State Department of Transportation stormwater characterization monitoring quality assurance project plan 2001/2003 Biennium: Olympia, Washington, Washington State Department of Transportation, 19 p.

Tetra Tech, Inc., 2003, Washington State Department of Transportation stormwater characterization/water quality management effectiveness monitoring 2003/2005 sampling season (rev. October 2003): Olympia, Washington, Washington State Department of Transportation, 31 p.

Tetra Tech, Inc., 2004, Washington State Department of Transportation stormwater characterization/water quality management effectiveness monitoring, addendum for 2004/2005 sampling season: Olympia, Washington, Washington State Department of Transportation, 16 p.

Thomson, N.R., McBean, E.A., and Mostrenko, I.B., 1996, Prediction and characterization of highway stormwater runoff quality: Ontario, Research and Development Branch, Ministry of Transportation, 98 p.

Thomson, N.R., McBean, E.A., Snodgrass, W., and Monstrenko, I.B., 1997, Highway stormwater runoff quality—Development of surrogate parameter relationships: Water, Air, and Soil Pollution, v. 94, p. 307–347.

Transportation Research Board, 1993, Research problem statements—Hydrology, hydraulics and water quality: Washington, D.C., Transportation Research Board, National Research Council Circular 405, p. 28–34.

Transportation Research Board, 1996a, 1996 Conference on environmental research needs in transportation—Resource papers and proposed research needs statements: Washington, D.C., Transportation Research Board, National Research Council Conference Proceedings, November 14–16, 277 p.

Transportation Research Board, 1996b, Research problem statements: Hydraulics, hydrology, and water quality: Washington, D.C., Transportation Research Board, National Research Council Circular 466, 39 p.

Transportation Research Board, 1997, Environmental research needs in transportation: Washington, D.C., Transportation Research Board, National Research Council Circular 469, 98 p.

Transportation Research Board, 2002, Environmental research needs in transportation: Washington D.C., Transportation Research Board, National Research Council, Conference Proceedings 28, 323 p.

U.S. Army Corps of Engineers, 1993, Hydrologic frequency analysis: U.S. Army Corps of Engineers Engineer Manual EM-1110-2-1415, 149 p.

U.S. Army Corps of Engineers, 1998, Evaluation of dredged material proposed for discharge in waters of the U.S.—Testing manual: Washington, D.C., U.S. Environmental Protection Agency, Report EPA 823-B-98-004, 113 p.

U.S. Environmental Protection Agency, 1992, NPDES storm water sampling guidance document: Washington, D.C., U.S. Environmental Protection Agency Technical Report EPA 833-B-92-001, 177 p.

U.S. Environmental Protection Agency, 1998, Guidance for data quality assessment—Practical methods for data analysis EPA QA/G-9: Washington, D.C., U.S. Environmental Protection Agency Report EPA/600/R-96/084, variously paged.

U.S. Environmental Protection Agency, 1999, Preliminary data summary of urban stormwater best management practices: Washington, D.C., U.S. Environmental Protection Agency Report EPA-821-R-99-012, 216 p.

U.S. Environmental Protection Agency, 2005a, National management measures to control nonpoint source pollution from urban areas: Washington, D.C., U.S. Environmental Protection Agency Report EPA 841-B-05-004, 518 p.

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

U.S. Geological Survey, 2000, Collection and use of total suspended solids data: Office of Water Quality and Office of Surface Water Technical Memorandum No. 2001.03, 3 p.

U.S. Geological Survey, 2005, The National Map—The Nation's topographic map for the 21st century: Accessed on the internet at URL June 15, 2005 at URL: http://nationalmap.gov/

U.S. Geological Survey, 2004, NWIS Web data for the Nation: Accessed December 24, 2006 at URL http://waterdata.usgs.gov.nwis/.

Venner, Marie; Leisenring, Marc; Pankani, Dan; and Strecker, Eric, 2004, Identification of research needs related to highway runoff management: Washington, D.C., Transportation Research Board, National Cooperative Highway Research Program NCHRP Report 521, 175 p.

Walsh, P.M., Barrett, M.E., Malina, J.F., Jr., and Charbeneau, R.J., 1997, Use of vegetative controls for treatment of highway runoff: Austin, TX, University of Texas at Austin, Bureau of Engineering Research, Center for Research in Water Resources Online Report 97- 5, 125 p.

Waschbusch, R.J., 2003, Data and methods of a 1999–2000 street sweeping study on an urban freeway in Milwaukee County, Wisconsin: U.S. Geological Survey Open-File Report 03-93, 41 p.

Washington State Department of Transportation, Environmental Services Office, 2001, 2001 NPDES progress report Cedar-Green, Island-Snohomish, and South Puget Sound water quality management areas, national pollutant discharge elimination system MS4 permits WASM10001, WASM20001, AND WASM30001: Olympia, Washington, Washington State Department of Transportation, 47 p.

Washington State Department of Transportation, Environmental Services Office, 2002, 2002 NPDES progress report Cedar-Green, Island-Snohomish, and South Puget Sound water quality management areas, national pollutant discharge elimination system MS4 permits WASM10001, WASM20001, AND WASM30001: Olympia, Washington, Washington State Department of Transportation, 57 p.

Washington State Department of Transportation, Environmental Services Office, 2003, 2003 NPDES progress report Cedar-Green, Island-Snohomish, and South Puget Sound water quality management areas, national pollutant discharge elimination system MS4 permits WASM10001, WASM20001, AND WASM30001: Olympia, Washington, Washington State Department of Transportation, 50 p.

Washington State Department of Transportation, Environmental Services Office, 2004, 2004 NPDES progress report Cedar-Green, Island-Snohomish, and South Puget Sound water quality management areas, national pollutant discharge elimination system MS4 permits WASM10001, WASM20001, AND WASM30001: Olympia, Washington, Washington State Department of Transportation, 797 p.

Weibull, W., 1939, The phenomenon of rupture in solids: Stockholm, Ingeniors Vetenskaps Akademien Handlinga 153, p. 17.

Appendix 1:

Application of the Robust Regression on Order Statistics (ROS) Method for Estimation of Summary Statistics for Data Sets with Values Below One or More Detection Limits

Proper statistical treatment of detection-limit data is important for interpretation of highway-runoff data (Driscoll and others, 1990; Strecker and others, 2001; Shumway and others, 2002) and other environmental data (Helsel and Cohn, 1988; Helsel, 2005). In highway and urban runoff studies, treatment of detection-limit data can profoundly affect interpretation of nutrient, trace element, and organic chemical data (Breault and Granato, 2003; Bricker, 2003; Lopes and Dionne, 2003). The U.S. Environmental Protection Agency (2003) defines a detection limit as the minimum concentration of an analyte (substance) that can be measured and reported with a 99-percent confidence that the analyte concentration is greater than zero. Values below such detection limits commonly are referred to as censored values, because the values are known only to be in the range from zero to the detection limit (Helsel, 2005). Detection limits for an analyte may vary with the analytical method used, may vary from laboratory to laboratory, and may vary over time for a given analytical method (Driscoll and others, 1990). Substitution methods, in which censored values are replaced with an arbitrary value (commonly the detection limit or one-half of the detection limit) are not statistically defensible, especially if a data set contains data with multiple detection limits (Helsel, 2005). The value of summary statistics calculated for these data sets depend on the arbitrary substitution value rather than information that is known about the rest of the data set. Therefore, a robust and defensible method is needed to produce planning-level estimates of population statistics for highway-runoff data.

Regression-on-order statistics (ROS) is a method that is considered robust for statistical estimation of summary statistics for data sets with censored values (Helsel, and Cohn, 1988; Shumway and others 2002; Helsel, 2005). Driscoll and others (1990) used the parametric version of the ROS method for estimating the standard deviation and the median of the data for estimating the mean for highway-runoff data in the 1990 Federal Highway Administration working database because highway-runoff data commonly are lognormally distributed. ROS is based on the solution of a regression equation formulated to predict the concentration of a sample based on its probability coordinate on a normal probability plot. In the fully parametric version, the slope and the intercept of the regression equation provide the estimate of the standard deviation and mean of the data set, respectively. This is referred to as the parametric version because it depends on the assumption that the whole data set is normally (or lognormally) distributed. Driscoll and others (1990) used the parametric version to determine the standard deviation in log-space, but used the median as an estimate of the geometric mean of the lognormal distribution. Driscoll and others (1990) used a nonparametric measure (the median) but they used the parametric assumption that the median equals the geometric mean, which is only quantitative if there is less than 50-percent censoring, and all censored values are less than the median value. The robust version of ROS is implemented by using the regression-line statistics to estimate a concentration for each censored value and by calculating summary statistics using standard methods with all estimated and measured values. This version of the ROS method is considered to be more robust because the assumption of (log) normality only applies to values estimated below one or more detection limits (Helsel and Cohn, 1988; Helsel, 2005). The ROS method is suitable for use when the amount of censoring is less than or equal to 80 percent of the data set (Helsel, 2005). The highway-runoff database application uses this robust ROS method.

The ROS method was implemented in the highway-runoff data set using Visual Basic for applications to query the underlying data set and to do the calculations necessary to estimate summary statistics. Any data point in the event mean concentration (EMC) value table (tblEMCValues) with a non-null EMC qualification code (field tEMCQual) is identified by the queries as a left-censored value below the nominal detection limit in the EMC field (sEMCValue). As stated in the body of the report, any value that should not be identified as a left-censored value should be qualified in the comment field (tEMCComment). A qualification-code editing form “frmFixQualCodes” is available in the Microsoft Access table window and can be used to identify and count EMC and qualification codes. Activation of this form also nullifies blank qualification codes by searching for blank fields (those with only one or more space-characters) and replacing them with null values. To estimate summary statistics, the user selects the event type(s), a runoff constituent, a data set or data sets, a monitoring site or sites, and a plotting-position formula to provide the information and data necessary for calculations. These statistics, however, are not recommended for use (Helsel, 2005).

Before doing the ROS estimate, the highway-runoff database application is designed to calculate summary statistics (including the average, standard deviation, skew, and median) by use of standard statistical formulas for data in linear space, natural log space, and log-base 10 space in a variety of ways (Helsel and Hirsch, 2002). Although omission of censored values and arbitrary censored-data-substitution methods are not recommended for rigorous statistical analysis of data (Helsel and Hirsch, 2002; Helsel, 2005), omission and substitution methods are used to show the range of summary statistics that these methods may produce. Substitution of values that are 1.0, 0.5, 0.1, 0.01, and 0.001 times the detection limit(s) are used in place of the censored values to calculate the summary statistics. Substitutions of zeroes for censored values are not included because highway-runoff data are assumed to approximate a lognormal distribution (Driscoll and others, 1990). The statistical estimates calculated with the ROS method should fall between estimates generated by substitution of values that are 1.0 and 0.001 times the detection limit(s)

If the number of censored values is greater than or equal to 50 percent, a censored median (Helsel and Hirsch, 2002) is calculated by the highway-runoff database application. For an odd number of points, this censored median is identified as being less than the mid-ranked data point. For an even number of points, this censored median is identified as being less than the larger of the two mid-ranked data points (rather than the average of the two mid-ranked data points as is done for uncensored data). The highway-runoff database application also calculates a "ROS" median irrespective of the censoring level using standard calculation methods with either data above detection limits or ROS estimates of censored values.

The highway-runoff database application calculates statistics by use of published methods (Helsel and Cohn, 1998; Helsel and others, 1988; Helsel, 2005). To calculate the ROS estimates, the censored and uncensored data values are ranked by a plotting-position formula (Hazen, 1914; Weibull, 1939; Blom, 1958; Gringorten, 1963; Cunnane, 1978). Helsel and Cohn (1998) indicate that the choice of plotting-position formula is of little consequence for estimates of the mean, standard deviation, and certain percentiles, but preliminary analysis with the highway-runoff data indicates that the Hazen plotting-position formula (Hazen, 1914; Helsel and Hirsch, 2002) may minimize calculated skew in the ROS estimates. Censored values for each detection limit are distributed evenly in the interval between the detection limit and zero (Helsel and Cohn, 1998; Helsel and others, 1988; Helsel, 2005). Once ranks are assigned to all censored and uncensored values and exceedence probabilities for these ranks are calculated, an estimate of the normal score for each probability is estimated. The normal score, commonly identified as the Z-score in statistical texts, is calculated from an empirical approximation to the normal cumulative distribution function (Abramowitz and Stegun, 1964, Equation 26.2.23). Uncensored values are transformed to their natural logarithmic values, and ordinary least-squares regression is used to estimate the equation of the regression line with the standard normal score as the explanatory variable. The estimated standard normal score for each censored value is used to estimate the natural log of the concentration value by use of the regression equation. The uncensored concentrations and censored estimates are then used to calculate summary statistics for retransformed values, for the natural log of concentration values, and for the log-base 10 of concentration values. The log-base 10 of concentration values are directly related by a factor of about 0.4343 times the natural log of concentration values, but are provided to facilitate use of the data.

There are a few technical issues, that should be considered for use of the ROS output. The number of significant figures reported do not reflect the uncertainty in input data, so calculated values should be rounded appropriately for use. The plotting-positions, standard normal score (Z-score), and individual estimated values are provided to facilitate examination, interpretation, and graphing of data. Also, the temporary table ttblMyROS has the unprocessed EMC values and qualification codes, the estimated EMC values, the ranked plotting-position values and the normal score for each EMC in the data set. These data may be exported but the values are deleted once the user exits the ROS form. Use of individual estimates for censored values, however, is not a recommended practice (Helsel, 2005). Helsel and Hirsch (2002) and Helsel (2005) provide guidance on how to graph data sets with estimated values.

References Cited

Abramowitz M., and Stegun, I. A. , 1964 Handbook of mathematical functions with formulas, graphs, and mathematical tables: Washington, D.C., National Bureau of Standards Applied Mathematics Series, v. 55, U.S. Government Printing Office, 1046 p.

Blom, G., 1958, Statistical estimates and transformed beta variables: New York, John Wiley, p. 68–75, 143–146.

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, v. 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. 165-234.

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, v. 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.

Cunnane, C., 1978, Unbiased plotting positions - a review: Journal of Hydrology, v. 37, p. 205–222.

Helsel, D.R., Cohn, T.A., and Harrington, John, 1988, Program to estimate summary statistics for data with multiple detection limits for Microsoft Windows (MDLWIN): Available on the internet at URL: http://www.practicalstats.com

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.

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, v. 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.

Shumway, R.H., Azari, R.S., Kayhanian, M., 2002, Statistical approaches to estimating mean water quality concentrations with detection limits: Environmental Science and Technology, v. 36, no. 15, p. 3345–3353.

Strecker, Eric; Mayo, Lynn; Quigley, Marcus; Howell, James, 2001, Guidance manual for monitoring highway runoff water quality: Washington, D.C., U.S. Department of Transportation, Federal Highway Administration, FHWA-EP-01-022, 206 p.

U.S. Environmental Protection Agency, 2003, Title 40—Protection of environment: Chapter I—Environmental Protection Agency Part 136—Guidelines establishing test procedures for the analysis of pollutants: Accessed on the internet on Aug 28, 2005 at URL http://www.access.gpo.gov/nara/cfr/waisidx_03/40cfr136_03.html

Weibull, W., 1939, The phenomenon of rupture in solids: Stockholm, Ingeniors Vetenskaps Akademien Handlinga 153, p. 17.

Granato and Cazenas—**Highway-Runoff Database (HRDB Version 1.0): A Data Warehouse and Preprocessor for the Stochastic Empirical Loading and Dilution Model**—FHWA-HEP-09-004